

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number  
**WO 02/070669 A2**

(51) International Patent Classification<sup>7</sup>: **C12N**

(21) International Application Number: PCT/US02/07719

(22) International Filing Date: 5 March 2002 (05.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/273,946	6 March 2001 (06.03.2001)	US
60/276,873	16 March 2001 (16.03.2001)	US
60/280,531	30 March 2001 (30.03.2001)	US
60/280,596	30 March 2001 (30.03.2001)	US
60/332,426	16 November 2001 (16.11.2001)	US
60/334,229	28 November 2001 (28.11.2001)	US
60/347,703	11 January 2002 (11.01.2002)	US

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **YANG, Jumming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polten Place Way, San Jose, CA 95121 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue #306, Sunnyvale, CA 94086 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). **LEE, Sally** [US/US]; 825 East Evelyn #425, Sunnyvale, CA 94086 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **GURURAJAN, Rajagopal** [IN/US]; 5591 Dent Avenue,

San Jose, CA 95118 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **KALLICK, Deborah, A.** [US/US]; 58 Linda Avenue, Atherton, CA 94027 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95014 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US).

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **SECRETED PROTEINS**

(57) Abstract: The invention provides human secreted proteins (SECP) and polynucleotides which identify and encode SECP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of SECP.

BEST AVAILABLE COPY

WO 02/070669 A2

## SECRETED PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of secreted proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of secreted proteins.

### BACKGROUND OF THE INVENTION

Protein transport and secretion are essential for cellular function. Protein transport is mediated by a signal peptide located at the amino terminus of the protein to be transported or secreted. The signal peptide is comprised of about ten to twenty hydrophobic amino acids which target the nascent protein from the ribosome to a particular membrane bound compartment such as the endoplasmic reticulum (ER). Proteins targeted to the ER may either proceed through the secretory pathway or remain in any of the secretory organelles such as the ER, Golgi apparatus, or lysosomes. Proteins that transit through the secretory pathway are either secreted into the extracellular space or retained in the plasma membrane. Proteins that are retained in the plasma membrane contain one or more transmembrane domains, each comprised of about 20 hydrophobic amino acid residues. Secreted proteins are generally synthesized as inactive precursors that are activated by post-translational processing events during transit through the secretory pathway. Such events include glycosylation, proteolysis, and removal of the signal peptide by a signal peptidase. Other events that may occur during protein transport include chaperone-dependent unfolding and folding of the nascent protein and interaction of the protein with a receptor or pore complex. Examples of secreted proteins with amino terminal signal peptides are discussed below and include proteins with important roles in cell-to-cell signaling. Such proteins include transmembrane receptors and cell surface markers, extracellular matrix molecules, cytokines, hormones, growth and differentiation factors, enzymes, neuropeptides, vasomediators, cell surface markers, and antigen recognition molecules. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of The Cell, Garland Publishing, New York, NY, pp. 557-560, 582-592.)

Cell surface markers include cell surface antigens identified on leukocytic cells of the immune system. These antigens have been identified using systematic, monoclonal antibody (mAb)-based "shot gun" techniques. These techniques have resulted in the production of hundreds of mAbs directed against unknown cell surface leukocytic antigens. These antigens have been grouped into "clusters of differentiation" based on common immunocytochemical localization patterns in various

differentiated and undifferentiated leukocytic cell types. Antigens in a given cluster are presumed to identify a single cell surface protein and are assigned a "cluster of differentiation" or "CD" designation. Some of the genes encoding proteins identified by CD antigens have been cloned and verified by standard molecular biology techniques. CD antigens have been characterized as both

5 transmembrane proteins and cell surface proteins anchored to the plasma membrane via covalent attachment to fatty acid-containing glycolipids such as glycosylphosphatidylinositol (GPI). (Reviewed in Barclay, A.N. et al. (1995) The Leucocyte Antigen Facts Book, Academic Press, San Diego, CA, pp. 17-20.)

Matrix proteins (MPs) are transmembrane and extracellular proteins which function in

10 formation, growth, remodeling, and maintenance of tissues and as important mediators and regulators of the inflammatory response. The expression and balance of MPs may be perturbed by biochemical changes that result from congenital, epigenetic, or infectious diseases. In addition, MPs affect leukocyte migration, proliferation, differentiation, and activation in the immune response. MPs are frequently characterized by the presence of one or more domains which may include collagen-like

15 domains, EGF-like domains, immunoglobulin-like domains, and fibronectin-like domains. In addition, MPs may be heavily glycosylated and may contain an Arginine-Glycine-Aspartate (RGD) tripeptide motif which may play a role in adhesive interactions. MPs include extracellular proteins such as fibronectin, collagen, galectin, vitronectin and its proteolytic derivative somatomedin B; and cell adhesion receptors such as cell adhesion molecules (CAMs), cadherins, and integrins. (Reviewed

20 in Ayad, S. et al. (1994) The Extracellular Matrix Facts Book, Academic Press, San Diego, CA, pp. 2-16; Ruoslahti, E. (1997) *Kidney Int.* 51:1413-1417; Sjaastad, M.D. and Nelson, W.J. (1997) *BioEssays* 19:47-55.)

Mucins are highly glycosylated glycoproteins that are the major structural component of the mucus gel. The physiological functions of mucins are cytoprotection, mechanical protection,

25 maintenance of viscosity in secretions, and cellular recognition. MUC6 is a human gastric mucin that is also found in gall bladder, pancreas, seminal vesicles, and female reproductive tract (Toribara, N.W. et al. (1997) *J. Biol. Chem.* 272:16398-16403). The MUC6 gene has been mapped to human chromosome 11 (Toribara, N.W. et al. (1993) *J. Biol. Chem.* 268:5879-5885). Hemomucin is a novel Drosophila surface mucin that may be involved in the induction of antibacterial effector molecules

30 (Theopold, U. et al. (1996) *J. Biol. Chem.* 271:12708-12715).

Tuftelins are one of four different enamel matrix proteins that have been identified so far. The other three known enamel matrix proteins are the amelogenins, enamelin and ameloblastin. Assembly of the enamel extracellular matrix from these component proteins is believed to be critical in producing a matrix competent to undergo mineral replacement. (Paine, C.T. et al. (1998) *Connect*

35 *Tissue Res.* 38:257-267). Tuftelin mRNA has been found to be expressed in human ameloblastoma

tumor, a non-mineralized odontogenic tumor (Deutsch, D. et al. (1998) Connect. Tissue Res. 39:177-184).

Olfactomedin-related proteins are extracellular matrix, secreted glycoproteins with conserved C-terminal motifs. They are expressed in a wide variety of tissues and in broad range of species, from *Caenorhabditis elegans* to *Homo sapiens*. Olfactomedin-related proteins comprise a gene family with at least 5 family members in humans. One of the five, TIGR/myocilin protein, is expressed in the eye and is associated with the pathogenesis of glaucoma (Kulkarni, N.H. et al. (2000) Genet. Res. 76:41-50). Research by Yokoyama et al. (1996) found a 135-amino acid protein, termed AMY, having 96% sequence identity with rat neuronal olfactomedin-related ER localized protein in a neuroblastoma cell line cDNA library, suggesting an essential role for AMY in nerve tissue (Yokoyama, M. et al. (1996) DNA Res. 3:311-320). Neuron-specific olfactomedin-related glycoproteins isolated from rat brain cDNA libraries show strong sequence similarity with olfactomedin. This similarity is suggestive of a matrix-related function of these glycoproteins in neurons and neurosecretory cells (Danielson, P.E. et al. (1994) J. Neurosci. Res. 38:468-478).

Mac-2 binding protein is a 90-kD serum protein (90K), a secreted glycoprotein isolated from both the human breast carcinoma cell line SK-BR-3, and human breast milk. It specifically binds to a human macrophage-associated lectin, Mac-2. Structurally, the mature protein is 567 amino acids in length and is preceded by an 18-amino acid leader. There are 16 cysteines and seven potential N-linked glycosylation sites. The first 106 amino acids represent a domain very similar to an ancient protein superfamily defined by a macrophage scavenger receptor cysteine-rich domain (Koths, K. et al. (1993) J. Biol. Chem. 268:14245-14249). 90K is elevated in the serum of subpopulations of AIDS patients and is expressed at varying levels in primary tumor samples and tumor cell lines. Ullrich et al. (1994) have demonstrated that 90K stimulates host defense systems and can induce interleukin-2 secretion. This immune stimulation is proposed to be a result of oncogenic transformation, viral infection or pathogenic invasion (Ullrich, A. et al. (1994) J. Biol. Chem. 269:18401-18407).

Semaphorins are a large group of axonal guidance molecules consisting of at least 30 different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth *in vitro*. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been suggested to have roles in protein-protein interactions and are thought to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94). Plexins are neuronal cell surface molecules that mediate cell adhesion via a homophilic binding mechanism in the presence of calcium ions. Plexins have been shown to be expressed in the receptors and neurons of



particular sensory systems (Ohta, K. et al. (1995) Cell 14:1189-1199). There is evidence that suggests that some plexins function to control motor and CNS axon guidance in the developing nervous system. Plexins, which themselves contain complete semaphorin domains, may be both the ancestors of classical semaphorins and binding partners for semaphorins (Winberg, M.L. et al (1998) 5 Cell 95:903-916).

Human pregnancy-specific beta 1-glycoprotein (PSG) is a family of closely related glycoproteins of molecular weights of 72 KDa, 64KDa, 62KDa, and 54KDa. Together with the carcinoembryonic antigen, they comprise a subfamily within the immunoglobulin superfamily (Plouzek, C.A. and Chou, J.Y. (1991) Endocrinology 129:950-958) Different subpopulations of PSG 10 have been found to be produced by the trophoblasts of the human placenta, and the amnionic and chorionic membranes (Plouzek, C.A. et al. (1993) Placenta 14:277-285).

Autocrine motility factor (AMF) is one of the motility cytokines regulating tumor cell migration; therefore identification of the signaling pathway coupled with it has critical importance. Autocrine motility factor receptor (AMFR) expression has been found to be associated with tumor 15 progression in thymoma (Ohta Y. et al. (2000) Int. J. Oncol. 17:259-264). AMFR is a cell surface glycoprotein of molecular weight 78KDa.

Hormones are secreted molecules that travel through the circulation and bind to specific receptors on the surface of, or within, target cells. Although they have diverse biochemical compositions and mechanisms of action, hormones can be grouped into two categories. One category 20 includes small lipophilic hormones that diffuse through the plasma membrane of target cells, bind to cytosolic or nuclear receptors, and form a complex that alters gene expression. Examples of these molecules include retinoic acid, thyroxine, and the cholesterol-derived steroid hormones such as progesterone, estrogen, testosterone, cortisol, and aldosterone. The second category includes hydrophilic hormones that function by binding to cell surface receptors that transduce signals across 25 the plasma membrane. Examples of such hormones include amino acid derivatives such as catecholamines (epinephrine, norepinephrine) and histamine, and peptide hormones such as glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotrophic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, and vasopressin. (See, for example, Lodish et al. (1995) Molecular Cell Biology, Scientific American Books Inc., New York, NY, pp. 856-864.)

30 Pro-opiomelanocortin (POMC) is the precursor polypeptide of corticotropin (ACTH), a hormone synthesized by the anterior pituitary gland, which functions in the stimulation of the adrenal cortex. POMC is also the precursor polypeptide of the hormone beta-lipotropin (beta-LPH). Each hormone includes smaller peptides with distinct biological activities: alpha-melanotropin (alpha-MSH) and corticotropin-like intermediate lobe peptide (CLIP) are formed from ACTH; gamma-lipotropin (gamma-LPH) and beta-endorphin are peptide components of beta-LPH; while beta-MSH 35

is contained within gamma-LPH. Adrenal insufficiency due to ACTH deficiency, resulting from a genetic mutation in exons 2 and 3 of POMC results in an endocrine disorder characterized by early-onset obesity, adrenal insufficiency, and red hair pigmentation (Chretien, M. et al. (1979) Can. J. Biochem. 57:1111-1121; Krude, H. et al. (1998) Nat. Genet. 19:155-157; Online Mendelian Inheritance in Man (OMIM) 176830).

Growth and differentiation factors are secreted proteins which function in intercellular communication. Some factors require oligomerization or association with membrane proteins for activity. Complex interactions among these factors and their receptors trigger intracellular signal transduction pathways that stimulate or inhibit cell division, cell differentiation, cell signaling, and cell motility. Most growth and differentiation factors act on cells in their local environment (paracrine signaling). There are three broad classes of growth and differentiation factors. The first class includes the large polypeptide growth factors such as epidermal growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, and platelet-derived growth factor. The second class includes the hematopoietic growth factors such as the colony stimulating factors (CSFs). Hematopoietic growth factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. The third class includes small peptide factors such as bombesin, vasopressin, oxytocin, endothelin, transferrin, angiotensin II, vasoactive intestinal peptide, and bradykinin, which function as hormones to regulate cellular functions other than proliferation.

Growth and differentiation factors play critical roles in neoplastic transformation of cells in vitro and in tumor progression in vivo. Inappropriate expression of growth factors by tumor cells may contribute to vascularization and metastasis of tumors. During hematopoiesis, growth factor misregulation can result in anemias, leukemias, and lymphomas. Certain growth factors such as interferon are cytotoxic to tumor cells both in vivo and in vitro. Moreover, some growth factors and growth factor receptors are related both structurally and functionally to oncoproteins. In addition, growth factors affect transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Reviewed in Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI, pp. 1-9.)

The Slit protein, first identified in *Drosophila*, is critical in central nervous system midline formation and potentially in nervous tissue histogenesis and axonal pathfinding. Itoh et al. ((1998) Brain Res. Mol. Brain Res. 62:175-186) have identified mammalian homologues of the slit gene (human Slit-1, Slit-2, Slit-3 and rat Slit-1). The encoded proteins are putative secreted proteins containing EGF-like motifs and leucine-rich repeats, both of which are conserved protein-protein interaction domains. Slit-1, -2, and -3 mRNAs are expressed in the brain, spinal cord, and thyroid, respectively (Itoh, A. et al., supra). The Slit family of proteins are indicated to be functional ligands of glypican-1 in nervous tissue and it is suggested that their interactions may be critical in certain

stages during central nervous system histogenesis (Liang, Y. et al. (1999) J. Biol. Chem. 274:17885-17892).

Neuropeptides and vasomediators (NP/VM) comprise a large family of endogenous signaling molecules. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related peptides involved in smooth muscle stimulation, vasopressin, vasoactive intestinal peptide, and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin and gastrin. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C.R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

NP/VMs are involved in numerous neurological and cardiovascular disorders. For example, neuropeptide Y is involved in hypertension, congestive heart failure, affective disorders, and appetite regulation. Somatostatin inhibits secretion of growth hormone and prolactin in the anterior pituitary, as well as inhibiting secretion in intestine, pancreatic acinar cells, and pancreatic beta-cells. A reduction in somatostatin levels has been reported in Alzheimer's disease and Parkinson's disease. Vasopressin acts in the kidney to increase water and sodium absorption, and in higher concentrations stimulates contraction of vascular smooth muscle, platelet activation, and glycogen breakdown in the liver. Vasopressin and its analogues are used clinically to treat diabetes insipidus. Endothelin and angiotensin are involved in hypertension, and drugs, such as captopril, which reduce plasma levels of angiotensin, are used to reduce blood pressure (Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 194; 252; 284; 55; 111).

Neuropeptides have also been shown to have roles in nociception (pain). Vasoactive intestinal peptide appears to play an important role in chronic neuropathic pain. Nociceptin, an endogenous ligand for the opioid receptor-like 1 receptor, is thought to have a predominantly anti-nociceptive effect, and has been shown to have analgesic properties in different animal models of tonic or chronic pain (Dickinson, T. and Fleetwood-Walker, S.M. (1998) Trends Pharmacol. Sci. 19:346-348).

Other proteins that contain signal peptides include secreted proteins with enzymatic activity. Such activity includes, for example, oxidoreductase/dehydrogenase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, or ligase activity. For example, matrix metalloproteinases are secreted hydrolytic enzymes that degrade the extracellular matrix and thus play an important role in tumor metastasis, tissue morphogenesis, and arthritis (Reponen, P. et al. (1995) Dev. Dyn. 202:388-396; Firestein, G.S. (1992) Curr. Opin. Rheumatol. 4:348-354; Ray, J.M.

and Stetler-Stevenson, W.G. (1994) Eur. Respir. J. 7:2062-2072; and Mignatti, P. and Rifkin, D.B. (1993) Physiol. Rev. 73:161-195). Lactate dehydrogenase A has been implicated in tumor induction by c-Myc (Lewis, B.C. et al. (2000) Cancer Res. 60:6178-6183). Additional examples are the acetyl-CoA synthetases which activate acetate for use in lipid synthesis or energy generation (Luong, A. et al. (2000) J. Biol. Chem. 275:26458-26466) and maspin, a serine protease inhibitor. The result of acetyl-CoA synthetase activity is the formation of acetyl-CoA from acetate and CoA. Acetyl-CoA synthetases share a region of sequence similarity identified as the AMP-binding domain signature. Acetyl-CoA synthetase has been shown to be associated with hypertension (Toh, H. (1991) Protein Seq. Data Anal. 4:111-117; and Iwai, N. et al. (1994) Hypertension 23:375-380). Maspin is related to the serpin family of protease inhibitors and has also been shown to play a role in the reversal of tumor progression, acting as a tumor suppressor. Maspin was identified in normal mammary epithelial cells, but was not identified in many mammary carcinoma cell lines, with the loss of expression most noticeable in advanced cancers. The use of maspin as a marker for tumor progression provides both a diagnostic and prognostic marker. Additionally, induction of maspin re-expression by pharmacological means may provide a promising therapeutic in the treatment of breast cancer (Zou, Z. et al. (1999) Science 263:526-529; Maass, N. et al. (2000) Acta Oncol. 39:931-934).

A number of isomerases catalyze steps in protein folding, phototransduction, and various anabolic and catabolic pathways. One class of isomerases is known as peptidyl-prolyl *cis-trans* isomerases (PPIases). PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. Two families of PPIases are the FK506 binding proteins (FKBPs), and cyclophilins (CyPs). FKBPs bind the potent immunosuppressants FK506 and rapamycin, thereby inhibiting signaling pathways in T-cells. Specifically, the PPIase activity of FKBPs is inhibited by binding of FK506 or rapamycin. There are five members of the FKBP family which are named according to their calculated molecular masses (FKBP12, FKBP13, FKBP25, FKBP52, and FKBP65), and localized to different regions of the cell where they associate with different protein complexes (Coss, M. et al. (1995) J. Biol. Chem. 270:29336-29341; Schreiber, S.L. (1991) Science 251:283-287).

The peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. CyP isomerase activity is associated with protein folding and protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the Hsp90/Hsc70 complex that binds steroid receptors. The mammalian CypA has been shown to bind the gag protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is

reversed by cyclosporin. This effect implicates CyPs in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) *J. Biol. Chem.* 266:23204-23214; Hunter, T. (1998) *Cell* 92:141-143; and Levenson, J.D. and Ness, S.A. (1998) *Mol. Cell.* 1:203-211).

Gamma-carboxyglutamic acid (Gla) proteins rich in proline (PRGPs) are members of a family of vitamin K-dependent single-pass integral membrane proteins. These proteins are characterized by an extracellular amino terminal domain of approximately 45 amino acids rich in Gla. The intracellular carboxyl terminal region contains one or two copies of the sequence PPXY, a motif present in a variety of proteins involved in such diverse cellular functions as signal transduction, cell cycle progression, and protein turnover (Kulman, J.D. et al. (2001) *Proc. Natl. Acad. Sci. USA* 98:1370-1375). The process of post-translational modification of glutamic residues to form Gla is Vitamin K-dependent carboxylation. Proteins which contain Gla include plasma proteins involved in blood coagulation. These proteins are prothrombin, proteins C, S, and Z, and coagulation factors VII, IX, and X. Osteocalcin (bone-Gla protein, BGP) and matrix Gla-protein (MGP) also contain Gla (Friedman, P.A. and C.T. Przysiecki (1987) *Int. J. Biochem.* 19:1-7; C. Vermeer (1990) *Biochem. J.* 266:625-636).

### Immunoglobulins

Antigen recognition molecules are key players in the sophisticated and complex immune systems which all vertebrates have developed to provide protection from viral, bacterial, fungal, and parasitic infections. A key feature of the immune system is its ability to distinguish foreign molecules, or antigens, from "self" molecules. This ability is mediated primarily by secreted and transmembrane proteins expressed by leukocytes (white blood cells) such as lymphocytes, granulocytes, and monocytes. Most of these proteins belong to the immunoglobulin (Ig) superfamily, members of which contain one or more repeats of a conserved structural domain. This Ig domain is comprised of antiparallel  $\beta$  sheets joined by a disulfide bond in an arrangement called the Ig fold. The criteria for a protein to be a member of the Ig superfamily is to have one or more Ig domains, which are regions of 70-110 amino acid residues in length homologous to either Ig variable-like (V) or Ig constant-like (C) domains. Members of the Ig superfamily include antibodies (Ab), T cell receptors (TCRs), class I and II major histocompatibility (MHC) proteins and immune cell-specific surface markers such as the "cluster of differentiation" or CD antigens, CD2, CD3, CD4, CD8, poly-Ig receptors, Fc receptors, neural cell-adhesion molecule (NCAM) and platelet-derived growth factor receptor (PDGFR).

Ig domains (V and C) are regions of conserved amino acid residues that give a polypeptide a globular tertiary structure called an immunoglobulin (or antibody) fold, which consists of two approximately parallel layers of  $\beta$ -sheets. Conserved cysteine residues form an intrachain disulfide-

bonded loop, 55-75 amino acid residues in length, which connects the two layers of  $\beta$ -sheets. Each  $\beta$ -sheet has three or four anti-parallel  $\beta$ -strands of 5-10 amino acid residues. Hydrophobic and hydrophilic interactions of amino acid residues within the  $\beta$ -strands stabilize the Ig fold (hydrophobic on inward facing amino acid residues and hydrophilic on the amino acid residues in the outward facing portion of the strands). A V domain consists of a longer polypeptide than a C domain, with an additional pair of  $\beta$ -strands in the Ig fold.

A consistent feature of Ig superfamily genes is that each sequence of an Ig domain is encoded by a single exon. It is possible that the superfamily evolved from a gene coding for a single Ig domain involved in mediating cell-cell interactions. New members of the superfamily then arose by exon and gene duplications. Modern Ig superfamily proteins contain different numbers of V and/or C domains. Another evolutionary feature of this superfamily is the ability to undergo DNA rearrangements, a unique feature retained by the antigen receptor members of the family.

Many members of the Ig superfamily are integral plasma membrane proteins with extracellular Ig domains. The hydrophobic amino acid residues of their transmembrane domains and their cytoplasmic tails are very diverse, with little or no homology among Ig family members or to known signal-transducing structures. There are exceptions to this general superfamily description. For example, the cytoplasmic tail of PDGFR has tyrosine kinase activity. In addition Thy-1 is a glycoprotein found on thymocytes and T cells. This protein has no cytoplasmic tail, but is instead attached to the plasma membrane by a covalent glycosphosphatidylinositol linkage.

Another common feature of many Ig superfamily proteins is the interactions between Ig domains which are essential for the function of these molecules. Interactions between Ig domains of a multimeric protein can be either homophilic or heterophilic (i.e., between the same or different Ig domains). Antibodies are multimeric proteins which have both homophilic and heterophilic interactions between Ig domains. Pairing of constant regions of heavy chains forms the Fc region of an antibody and pairing of variable regions of light and heavy chains form the antigen binding site of an antibody. Heterophilic interactions also occur between Ig domains of different molecules. These interactions provide adhesion between cells for significant cell-cell interactions in the immune system and in the developing and mature nervous system. (Reviewed in Abbas, A.K. et al. (1991) Cellular and Molecular Immunology, W.B. Saunders Company, Philadelphia, PA, pp.142-145.)

### Antibodies

MHC proteins are cell surface markers that bind to and present foreign antigens to T cells. MHC molecules are classified as either class I or class II. Class I MHC molecules (MHC I) are expressed on the surface of almost all cells and are involved in the presentation of antigen to cytotoxic T cells. For example, a cell infected with virus will degrade intracellular viral proteins and express the protein fragments bound to MHC I molecules on the cell surface. The MHC I/antigen

complex is recognized by cytotoxic T-cells which destroy the infected cell and the virus within. Class II MHC molecules are expressed primarily on specialized antigen-presenting cells of the immune system, such as B-cells and macrophages. These cells ingest foreign proteins from the extracellular fluid and express MHC II/antigen complex on the cell surface. This complex activates helper T-cells, which then secrete cytokines and other factors that stimulate the immune response. MHC molecules also play an important role in organ rejection following transplantation. Rejection occurs when the recipient's T-cells respond to foreign MHC molecules on the transplanted organ in the same way as to self MHC molecules bound to foreign antigen. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York, NY, pp. 1229-1246.)

Antibodies are multimeric members of the Ig superfamily which are either expressed on the surface of B-cells or secreted by B-cells into the circulation. Antibodies bind and neutralize foreign antigens in the blood and other extracellular fluids. The prototypical antibody is a tetramer consisting of two identical heavy polypeptide chains (H-chains) and two identical light polypeptide chains (L-chains) interlinked by disulfide bonds. This arrangement confers the characteristic Y-shape to antibody molecules. Antibodies are classified based on their H-chain composition. The five antibody classes, IgA, IgD, IgE, IgG and IgM, are defined by the  $\alpha$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma$ , and  $\mu$  H-chain types. There are two types of L-chains,  $\kappa$  and  $\lambda$ , either of which may associate as a pair with any H-chain pair. IgG, the most common class of antibody found in the circulation, is tetrameric, while the other classes of antibodies are generally variants or multimers of this basic structure.

H-chains and L-chains each contain an N-terminal variable region and a C-terminal constant region. The constant region consists of about 110 amino acids in L-chains and about 330 or 440 amino acids in H-chains. The amino acid sequence of the constant region is nearly identical among H- or L-chains of a particular class. The variable region consists of about 110 amino acids in both H- and L-chains. However, the amino acid sequence of the variable region differs among H- or L-chains of a particular class. Within each H- or L-chain variable region are three hypervariable regions of extensive sequence diversity, each consisting of about 5 to 10 amino acids. In the antibody molecule, the H- and L-chain hypervariable regions come together to form the antigen recognition site. (Reviewed in Alberts, B. et al. supra, pp. 1206-1213 and 1216-1217.)

Both H-chains and L-chains contain the repeated Ig domains of members of the Ig superfamily. For example, a typical H-chain contains four Ig domains, three of which occur within the constant region and one of which occurs within the variable region and contributes to the formation of the antigen recognition site. Likewise, a typical L-chain contains two Ig domains, one of which occurs within the constant region and one of which occurs within the variable region.

The immune system is capable of recognizing and responding to any foreign molecule that enters the body. Therefore, the immune system must be armed with a full repertoire of antibodies

against all potential antigens. Such antibody diversity is generated by somatic rearrangement of gene segments encoding variable and constant regions. These gene segments are joined together by site-specific recombination which occurs between highly conserved DNA sequences that flank each gene segment. Because there are hundreds of different gene segments, millions of unique genes can be generated combinatorially. In addition, imprecise joining of these segments and an unusually high rate of somatic mutation within these segments further contribute to the generation of a diverse antibody population.

#### Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The discovery of new secreted proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of secreted proteins.

### **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, secreted proteins, referred to collectively as "SECP" and individually as "SECP-1," "SECP-2," "SECP-3," "SECP-4," "SECP-5," "SECP-6," "SECP-7," "SECP-8," "SECP-9," "SECP-10," "SECP-11," "SECP-12," "SECP-13," "SECP-14," "SECP-15," "SECP-16," "SECP-17," "SECP-18," "SECP-19," "SECP-20," "SECP-21," "SECP-22," and "SECP-23." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-23.



The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-23. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:24-46.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a

polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of

treating a disease or condition associated with overexpression of functional SECP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20

contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a

5 polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID  
10 NO:24-46, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting  
15 of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

## 20 BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME  
25 database homologs, for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

30 Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of  
35 the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"SECP" refers to the amino acid sequences of substantially purified SECP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of SECP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of SECP either by directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

An "allelic variant" is an alternative form of the gene encoding SECP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to

allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding SECP include those sequences with deletions, 5 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as SECP or a polypeptide with at least one functional characteristic of SECP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding SECP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding 10 SECP. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent SECP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of SECP is retained. For example, 15 negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

20 The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

25 “Amplification” relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of SECP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small 30 molecules, or any other compound or composition which modulates the activity of SECP either by directly interacting with SECP or by acting on components of the biological pathway in which SECP participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. 35 Antibodies that bind SECP polypeptides can be prepared using intact polypeptides or using fragments

containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having



modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic SECP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding SECP or fragments of SECP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
5	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
10	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
15	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
20	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the

evolution of new protein functions.

A "fragment" is a unique portion of SECP or the polynucleotide encoding SECP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:24-46 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:24-46, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:24-46 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:24-46 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:24-46 and the region of SEQ ID NO:24-46 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-23 is encoded by a fragment of SEQ ID NO:24-46. A fragment of SEQ ID NO:1-23 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-23. For example, a fragment of SEQ ID NO:1-23 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-23. The precise length of a fragment of SEQ ID NO:1-23 and the region of SEQ ID NO:1-23 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps

in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

25        *Matrix: BLOSUM62*  
          *Reward for match: 1*  
          *Penalty for mismatch: -2*  
          *Open Gap: 5 and Extension Gap: 2 penalties*  
          *Gap x drop-off: 50*  
 30        *Expect: 10*  
          *Word Size: 11*  
          *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,  
 35 over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at

least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

5 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

10 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

15 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default  
20 residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for  
25 example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

30 *Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for  
35 instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least

150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain  
5 DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

10 “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining  
15 stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity.  
20 Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic  
25 strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

30 High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents  
35 include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic

solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of SECP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of SECP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of SECP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of SECP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably

linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

“Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

“Post-translational modification” of an SECP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of SECP.

“Probe” refers to nucleic acid sequences encoding SECP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

“Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such



(UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term “sample” is used in its broadest sense. A sample suspected of containing SECP, nucleic acids encoding SECP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A “transcript image” or “expression profile” refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

“Transformation” describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods

purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the

5 PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which

10 sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments,

15 thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to

20 identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the

25 artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to

30 transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated

35 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions

well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term

5 “transformed cells” includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A “transgenic organism,” as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene

between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

- 5 A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least  
10 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

- The invention is based on the discovery of new human secreted proteins (SECP), the  
15 polynucleotides encoding SECP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders.

- Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a  
20 single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column  
25 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

- Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database.  
30 Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the  
35 matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the

GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs, including the locations of signal peptides (as indicated by "Signal Peptide" and/or "signal\_cleavage".) Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are secreted proteins. For example, SEQ ID NO:1 is 91% identical, from residue M1 to residue L371, to human lactate dehydrogenase A (GenBank ID g12331000) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $1.2e-180$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a lactate/malate dehydrogenase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a lactate dehydrogenase.

In another example, SEQ ID NO:8 is 32% identical, from residue S80 to residue A268, to human Slit-1 protein (GenBank ID g4049585) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $1.3e-19$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains leucine rich repeats, a leucine rich repeat C-terminal domain and a leucine rich repeat N-terminal domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:8 is a secreted protein (note that Slit proteins encode putative secreted proteins, which contain among other motifs, leucine-rich repeats).

In another example, SEQ ID NO:9 is 100% identical, from residue K9 to residue V356, to human bA425Ab.2 (similar to connexin) (GenBank ID g10334641) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $2.6e-190$ , which

indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:9 also contains a connexin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:9 is a connexin containing protein (note that “connexins” are hexamers of integral membrane proteins which make up connexons, the closely packed pairs of transmembrane channels which make up gap junctions through which small molecules diffuse between cells).

In a further example, SEQ ID NO:19 contains signal peptide domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provide further corroborative evidence that SEQ ID NO:19 is a secreted protein. SEQ ID NO:2-7, SEQ ID NO:10-18 and SEQ ID NO:20-23 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-23 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:24-46 or that distinguish between SEQ ID NO:24-46 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation “ENST”). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation “NP”). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by

an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3,4</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (i.e., gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors

which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses SECP variants. A preferred SECP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the SECP amino acid sequence, and which contains at least one functional or structural characteristic of SECP.

The invention also encompasses polynucleotides which encode SECP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46, which encodes SECP. The polynucleotide sequences of SEQ ID NO:24-46, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding SECP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding SECP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:24-46 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:24-46. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of SECP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding SECP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding SECP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding SECP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding SECP. For example, a polynucleotide comprising a sequence of SEQ ID NO:44 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:37, a polynucleotide comprising a sequence of SEQ ID NO:45 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:38, and a polynucleotide comprising a sequence of SEQ ID NO:46 is a splice variant of a polynucleotide comprising a sequence of SEQ ID NO:43. Any one of the splice variants described above can encode an amino



acid sequence which contains at least one functional or structural characteristic of SECP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding SECP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring SECP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode SECP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring SECP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding SECP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding SECP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode SECP and SECP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding SECP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:24-46 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE

amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding SECP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode SECP may be cloned in recombinant DNA molecules that direct expression of SECP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express SECP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter SECP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of SECP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby

maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding SECP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.)

Alternatively, SECP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of SECP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active SECP, the nucleotide sequences encoding SECP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding SECP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding SECP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding SECP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding SECP and appropriate transcriptional and translational control

elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

5 A variety of expression vector/host systems may be utilized to contain and express sequences encoding SECP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV,  
10 or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New  
15 York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci.*  
20 *USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding SECP. For example, routine cloning,  
25 subcloning, and propagation of polynucleotide sequences encoding SECP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding SECP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for  
30 in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of SECP are needed, e.g. for the production of antibodies, vectors which direct high level expression of SECP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

35 Yeast expression systems may be used for production of SECP. A number of vectors

containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 5 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of SECP. Transcription of sequences encoding SECP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 10 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology 15 (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding SECP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain 20 infective virus which expresses SECP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of 25 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression 30 of SECP in cell lines is preferred. For example, sequences encoding SECP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance 35 to a selective agent, and its presence allows growth and recovery of cells which successfully express

e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

5 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding SECP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding SECP, or any fragments thereof, may be cloned into a vector  
10 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for  
15 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding SECP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence  
20 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode SECP may be designed to contain signal sequences which direct secretion of SECP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of  
25 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the  
30 American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding SECP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric SECP protein  
35 containing a heterologous moiety that can be recognized by a commercially available antibody may

the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk<sup>-</sup>* and *apr<sup>-</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; Lowy, I. et al. (1980) *Cell* 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding SECP is inserted within a marker gene sequence, transformed cells containing sequences encoding SECP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding SECP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding SECP and that express SECP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of SECP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on SECP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,



facilitate the screening of peptide libraries for inhibitors of SECP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the SECP encoding sequence and the heterologous protein sequence, so that SECP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled SECP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

SECP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to SECP. At least one and up to a plurality of test compounds may be screened for specific binding to SECP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of SECP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which SECP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express SECP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing SECP or cell membrane fractions which contain SECP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either SECP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with SECP, either in solution or affixed to a solid support, and detecting the binding of SECP to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a  
5 labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

SECP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of SECP. Such compounds may include agonists, antagonists, or partial or  
10 inverse agonists. In one embodiment, an assay is performed under conditions permissive for SECP activity, wherein SECP is combined with at least one test compound, and the activity of SECP in the presence of a test compound is compared with the activity of SECP in the absence of the test compound. A change in the activity of SECP in the presence of the test compound is indicative of a compound that modulates the activity of SECP. Alternatively, a test compound is combined with an  
15 in vitro or cell-free system comprising SECP under conditions suitable for SECP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of SECP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding SECP or their mammalian homologs may  
20 be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of  
25 interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids  
30 Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding SECP may also be manipulated in vitro in ES cells derived from

human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

- 5 Polynucleotides encoding SECP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding SECP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and
- 10 treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress SECP, e.g., by secreting SECP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

### **THERAPEUTICS**

- Chemical and structural similarity, e.g., in the context of sequences and motifs, exists
- 15 between regions of SECP and secreted proteins. In addition, examples of tissues expressing SECP are brain, cardiac, and lung tissues, prostate, adrenal, rectal and ovarian tumors, digestive, reproductive and testicular tissues tissue, neurological tissue, cardiovascular tissue, urological tissue, cancerous lung tissue, and can also be found in Table 6. Therefore, SECP appears to play a role in cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental
- 20 disorders. In the treatment of disorders associated with increased SECP expression or activity, it is desirable to decrease the expression or activity of SECP. In the treatment of disorders associated with decreased SECP expression or activity, it is desirable to increase the expression or activity of SECP.

- Therefore, in one embodiment, SECP or a fragment or derivative thereof may be
- 25 administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma,
- 30 leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory
- 35 distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis,

autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,

5 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner

10 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart

15 disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis,

20 balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating

25 diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

30 encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

35 disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD);

akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss.

In another embodiment, a vector capable of expressing SECP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified SECP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of SECP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of SECP including, but not limited to, those listed above.

In a further embodiment, an antagonist of SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, cardiovascular, neurological, and developmental disorders described above. In one aspect, an antibody which specifically binds SECP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express SECP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding SECP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of SECP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the

various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of SECP may be produced using methods which are generally known in the art.

In particular, purified SECP may be used to produce antibodies or to screen libraries of

5 pharmaceutical agents to identify those which specifically bind SECP. Antibodies to SECP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from  
10 camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with SECP or with any  
15 fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are  
20 especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to SECP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches  
25 of SECP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to SECP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma  
30 technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate  
35 antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc.

Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce SECP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for SECP may also be generated. For example, such fragments include, but are not limited to,  $F(ab')_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between SECP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering SECP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for SECP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of SECP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple SECP epitopes, represents the average affinity, or avidity, of the antibodies for SECP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular SECP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the SECP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of SECP, preferably in active form, from the

antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of SECP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding SECP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding SECP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding SECP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding SECP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475),



cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in SECP expression or regulation causes disease, the expression of SECP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in SECP are treated by constructing mammalian expression vectors encoding SECP and introducing these vectors by mechanical means into SECP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of SECP include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). SECP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding SECP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to SECP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding SECP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding SECP to cells which have one or more genetic abnormalities with respect to the expression of SECP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999)

Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding SECP to target cells which have one or more genetic abnormalities with respect to the expression of SECP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing SECP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding SECP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for SECP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of SECP-coding RNAs and the synthesis of high levels of SECP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will

allow the introduction of SECP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding SECP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding SECP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding SECP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased SECP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding SECP may be therapeutically useful, and in the treatment of disorders associated with decreased SECP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding SECP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding SECP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding SECP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding SECP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the

polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res.

Commun. 268:8-13). A particular embodiment of the present invention involves screening a  
5 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

10 Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat.

15 Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition  
20 which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of SECP, antibodies to SECP, and mimetics, agonists, antagonists, or inhibitors of SECP.

25 The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form.  
30 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton,  
35 J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration

without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

5 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising SECP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, SECP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to  
10 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration  
15 range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example SECP or fragments thereof, antibodies of SECP, and agonists, antagonists or inhibitors of SECP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by  
20 standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are  
25 used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the  
30 subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week,  
35 or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind SECP may be used for the diagnosis of disorders characterized by expression of SECP, or in assays to monitor patients being treated with SECP or agonists, antagonists, or inhibitors of SECP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for SECP include methods which utilize the antibody and a label to detect SECP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring SECP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of SECP expression. Normal or standard values for SECP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to SECP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of SECP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding SECP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of SECP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of SECP, and to monitor regulation of SECP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding SECP or closely related molecules may be used to identify nucleic acid sequences which encode SECP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding SECP, allelic variants, or related



sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the SECP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:24-46 or from genomic sequences including promoters, enhancers, and introns of the SECP gene.

Means for producing specific hybridization probes for DNAs encoding SECP include the cloning of polynucleotide sequences encoding SECP or SECP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding SECP may be used for the diagnosis of disorders associated with expression of SECP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as congestive heart failure, ischemic heart disease, angina pectoris,

myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss. The polynucleotide sequences encoding SECP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and

multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered SECP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding SECP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding SECP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding SECP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of SECP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding SECP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding SECP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide

encoding SECP, or a fragment of a polynucleotide complementary to the polynucleotide encoding SECP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

5 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding SECP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP,  
10 oligonucleotide primers derived from the polynucleotide sequences encoding SECP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the  
15 oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation  
20 of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also  
25 useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in  
30 N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations  
35 and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu

(1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of SECP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, SECP, fragments of SECP, or antibodies specific for SECP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines,

biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by

separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for SECP to quantify the levels of SECP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological

sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

5 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of  
10 protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al.  
15 (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding SECP may be used  
20 to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a  
25 chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial PI constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop  
30 genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic  
35 map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man



(OMIM) World Wide Web site. Correlation between the location of the gene encoding SECP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, SECP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between SECP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with SECP, or fragments thereof, and washed. Bound SECP is then detected by methods well known in the art. Purified SECP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding SECP specifically compete with a test compound for binding SECP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with SECP.

In additional embodiments, the nucleotide sequences which encode SECP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

5 The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/280,531, U.S. Ser. No. 60/280,596, U.S. Ser. No. 60/276,873, U.S. Ser. No. 60/273,946, U.S. Ser. No. 60/332,426, U.S. Ser. No. 60/334,229 and U.S. Ser. No. 60/347,703, are expressly incorporated by reference herein.

### EXAMPLES

#### 10 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine  
15 isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated  
20 using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA  
25 libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the  
30 appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSORT1 plasmid (Life Technologies), PCDNA2.1 plasmid  
35 (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen),

PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## 5 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, 10 QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4 °C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal 15 cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

20 Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared 25 using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI 30 protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing 35 vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and

programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens,

5 Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) 10 Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, 15 or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention 20 may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full 25 length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

30 Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column 35 presents, where applicable, the scores, probability values, and other parameters used to evaluate the

strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:24-46. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative secreted proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode secreted proteins, the encoded polypeptides were analyzed by querying against PFAM models for secreted proteins. Potential secreted proteins were also identified by homology to Incyte cDNA sequences that had been annotated as secreted proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information,

generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of SECP Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:24-46 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:24-46 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:37 was mapped to chromosome 5 within the interval from 134.90 to 141.40 centiMorgans.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate

the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding SECP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding SECP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### **VIII. Extension of SECP Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ ,



and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the  
5 alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE  
10 and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the  
15 sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%)  
20 agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in  
25 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was  
30 quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

35 In like manner, full length polynucleotide sequences are verified using the above procedure or

are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Identification of Single Nucleotide Polymorphisms in SECP Encoding Polynucleotides**

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:24-46 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

#### **X. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:24-46 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a

SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

5 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and  
10 compared.

#### **XI. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the  
15 aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to  
20 those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be  
25 selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser  
30 desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **Tissue or Cell Sample Preparation**

35 Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and

poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### 15 Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in

0.2% SDS and distilled water as before.

### **Hybridization**

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

### **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital

(A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## **XII. Complementary Polynucleotides**

Sequences complementary to the SECP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring SECP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of SECP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the SECP-encoding transcript.

## **XIII. Expression of SECP**

Expression and purification of SECP is achieved using bacterial or virus-based expression systems. For expression of SECP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express SECP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of SECP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding SECP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K.

et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, SECP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from SECP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified SECP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII and XIX where applicable.

#### **XIV. Functional Assays**

SECP function is assessed by expressing the sequences encoding SECP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are

discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of SECP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding SECP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding SECP and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### 10 XV. Production of SECP Specific Antibodies

SECP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the SECP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer. (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-SECP activity by, for example, binding the peptide or SECP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XVI. Purification of Naturally Occurring SECP Using Specific Antibodies

Naturally occurring or recombinant SECP is substantially purified by immunoaffinity chromatography using antibodies specific for SECP. An immunoaffinity column is constructed by covalently coupling anti-SECP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing SECP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of SECP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt



antibody/SECP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and SECP is collected.

#### **XVII. Identification of Molecules Which Interact with SECP**

SECP, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent.

5 (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled SECP, washed, and any wells with labeled SECP complex are assayed. Data obtained using different concentrations of SECP are used to calculate values for the number, affinity, and association of SECP with the candidate molecules.

10 Alternatively, molecules interacting with SECP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

SECP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions  
15 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### **XVIII. Demonstration of SECP Activity**

An assay for growth stimulating or inhibiting activity of SECP measures the amount of DNA synthesis in Swiss mouse 3T3 cells (McKay, I. and Leigh, I., eds. (1993) Growth Factors: A Practical  
20 Approach, Oxford University Press, New York, NY). In this assay, varying amounts of SECP are added to quiescent 3T3 cultured cells in the presence of [ $^3\text{H}$ ]thymidine, a radioactive DNA precursor. SECP for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [ $^3\text{H}$ ]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized  
25 DNA. A linear dose-response curve over at least a hundred-fold SECP concentration range is indicative of growth modulating activity. One unit of activity per milliliter is defined as the concentration of SECP producing a 50% response level, where 100% represents maximal incorporation of [ $^3\text{H}$ ]thymidine into acid-precipitable DNA.

Alternatively, an assay for SECP activity measures the stimulation or inhibition of  
30 neurotransmission in cultured cells. Cultured CHO fibroblasts are exposed to SECP. Following endocytic uptake of SECP, the cells are washed with fresh culture medium, and a whole cell voltage-clamped Xenopus myocyte is manipulated into contact with one of the fibroblasts in SECP-free medium. Membrane currents are recorded from the myocyte. Increased or decreased current relative to control values are indicative of neuromodulatory effects of SECP (Morimoto, T. et al. (1995)  
35 *Neuron* 15:689-696).

Alternatively, an assay for SECP activity measures the amount of SECP in secretory, membrane-bound organelles. Transfected cells as described above are harvested and lysed. The lysate is fractionated using methods known to those of skill in the art, for example, sucrose gradient ultracentrifugation. Such methods allow the isolation of subcellular components such as the Golgi apparatus, ER, small membrane-bound vesicles, and other secretory organelles. Immunoprecipitations from fractionated and total cell lysates are performed using SECP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The concentration of SECP in secretory organelles relative to SECP in total cell lysate is proportional to the amount of SECP in transit through the secretory pathway.

Alternatively, AMP binding activity is measured by combining SECP with  $^{32}\text{P}$ -labeled AMP. The reaction is incubated at  $37^\circ\text{C}$  and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to remove unbound label. The radioactivity retained in the gel is proportional to SECP activity.

A microtubule motility assay for SECP measures motor protein activity. In this assay, recombinant SECP is immobilized onto a glass slide or similar substrate. Taxol-stabilized bovine brain microtubules (commercially available) in a solution containing ATP and cytosolic extract are perfused onto the slide. Movement of microtubules as driven by SECP motor activity can be visualized and quantified using video-enhanced light microscopy and image analysis techniques. SECP activity is directly proportional to the frequency and velocity of microtubule movement.

Alternatively, an assay for SECP measures the formation of protein filaments in vitro. A solution of SECP at a concentration greater than the "critical concentration" for polymer assembly is applied to carbon-coated grids. Appropriate nucleation sites may be supplied in the solution. The grids are negative stained with 0.7% (w/v) aqueous uranyl acetate and examined by electron microscopy. The appearance of filaments of approximately 25 nm (microtubules), 8 nm (actin), or 10 nm (intermediate filaments) is a demonstration of SECP activity.

In another alternative, SECP activity is measured by the binding of SECP to protein filaments.  $^{35}\text{S}$ -Met labeled SECP sample is incubated with the appropriate filament protein (actin, tubulin, or intermediate filament protein) and complexed protein is collected by immunoprecipitation using an antibody against the filament protein. The immunoprecipitate is then run out on SDS-PAGE and the amount of SECP bound is measured by autoradiography.

#### **XIX. Demonstration of Immunoglobulin Activity**

An assay for SECP activity measures the ability of SECP to recognize and precipitate antigens from serum. This activity can be measured by the quantitative precipitin reaction. (Golub, E.S. et al. (1987) Immunology: A Synthesis, Sinauer Associates, Sunderland, MA, pages 113-115.)

SECP is isotopically labeled using methods known in the art. Various serum concentrations are

added to constant amounts of labeled SECP. SECP-antigen complexes precipitate out of solution and are collected by centrifugation. The amount of precipitable SECP-antigen complex is proportional to the amount of radioisotope detected in the precipitate. The amount of precipitable SECP-antigen complex is plotted against the serum concentration. For various serum concentrations, a  
5 characteristic precipitin curve is obtained, in which the amount of precipitable SECP-antigen complex initially increases proportionately with increasing serum concentration, peaks at the equivalence point, and then decreases proportionately with further increases in serum concentration. Thus, the amount of precipitable SECP-antigen complex is a measure of SECP activity which is characterized by sensitivity to both limiting and excess quantities of antigen.

10 Alternatively, an assay for SECP activity measures the expression of SECP on the cell surface. cDNA encoding SECP is transfected into a non-leukocytic cell line. Cell surface proteins are labeled with biotin (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using SECP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled  
15 immunoprecipitant is proportional to the amount of SECP expressed on the cell surface.

Alternatively, an assay for SECP activity measures the amount of cell aggregation induced by overexpression of SECP. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding SECP contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green  
20 Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of SECP activity.

Various modifications and variations of the described methods and systems of the invention  
25 will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the  
30 scope of the following claims.

Table 1

IncYTE Project ID	Polypeptide SEQ ID NO:	IncYTE Polypeptide ID	Polynucleotide SEQ ID NO:	IncYTE Polynucleotide ID	CA2 Reagents
6024712	1	6024712CD1	24	6024712CB1	
72176922	2	72176922CD1	25	72176922CB1	1840186CA2, 656258CA2, 90108583CA2
1392717	3	1392717CD1	26	1392717CB1	1392717CA2, 90066808CA2, 90066907CA2, 90066915CA2, 90066923CA2, 90066939CA2, 90067015CA2, 90067031CA2, 90067063CA2
2701254	4	2701254CD1	27	2701254CB1	5944001CA2
71774318	5	71774318CD1	28	71774318CB1	90067016CA2
71802522	6	71802522CD1	29	71802522CB1	3068613CA2
6425956	7	6425956CD1	30	6425956CB1	90092901CA2, 90092925CA2
7494288	8	7494288CD1	31	7494288CB1	90078552CA2, 90078560CA2, 90078576CA2, 90078584CA2
7474330	9	7474330CD1	32	7474330CB1	90055956CA2, 90055980CA2, 90055996CA2, 90056064CA2, 90056096CA2
5911370	10	5911370CD1	33	5911370CB1	6269343CA2, 6269670CA2

Table 1

IncYTE Project ID	Polypeptide SEQ ID NO:	IncYTE Polypeptide ID	Polynucleotide SEQ ID NO:	IncYTE Polynucleotide ID	CA2 Reagents
7647134	11	7647134CD1	34	7647134CB1	
1631327	12	1631327CD1	35	1631327CB1	
44232	13	044232CD1	36	044232CB1	2169223CA2
560293	14	560293CD1	37	560293CB1	90059273CA2
2025618	15	2025618CD1	38	2025618CB1	
3342443	16	3342443CD1	39	3342443CB1	
2267957	17	2267957CD1	40	2267957CB1	90080362CA2, 90080370CA2, 90080394CA2, 90080462CA2, 90080470CA2, 90080478CA2
7480277	18	7480277CD1	41	7480277CB1	
3450647	19	3450647CD1	42	3450647CB1	3450647CA2
2053428	20	2053428CD1	43	2053428CB1	
7503614	21	7503614CD1	44	7503614CB1	
7503456	22	7503456CD1	45	7503456CB1	
7503459	23	7503459CD1	46	7503459CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	6024712CD1	g12331000	1.20E-180	[Homo sapiens] lactate dehydrogenase A
2	72176922CD1	g6563042	4.00E-13	[Homo sapiens] leukocyte-associated Ig-like receptor 1b
8	7494288CD1	g4049585	1.30E-19	[Homo sapiens] Slit-1 protein
9	7474330CD1	g15990853	0	Itoh, A. et al., (1998) Brain Res. Mol. Brain Res. 62 (2), 175-186
10	5911370CD1	g790641	2.00E-25	[Homo sapiens] connexin40.1
11	7647134CD1	g1469415	1.20E-215	[Hordeum vulgare] gamma-thionin
12	1631327CD1	g13507259	0	[Homo sapiens] paired-box protein PAX2
15	2025618CD1	g2947228	4.30E-16	Sanyanusin, P. (1995) Nat. Genet. 9:358-364
17	2267957CD1	g532493	5.40E-21	[Homo sapiens] amnionless
18	7480277CD1	g7529598	3.30E-119	[Plasmodium yoelii yoelii] erythrocyte binding protein
22	7503456CD1	g3549261	1.9E-12	[Mus musculus] SLIT1
22	7503456CD1	252694 Y57G11 C.20	1.8E-11	[Homo sapiens] d1402N21.3 (novel protein with Immunoglobulin domains)
22	7503456CD1	623900 MYH3	1.70E-10	[Dictyostelium discoideum] interaptin
23	7503459CD1	g13872536	1.60E-12	Rivero, F. et al., J. Cell Biol. 142:735-750 (1998)
23	7503459CD1	248546 R02D3.2	5.5E-20	[Caenorhabditis elegans] Putative paralog of C. elegans K09F6.6, has similarity to C. elegans NMY-1, a myosin family member
				[Homo sapiens][Motor protein; Hydrolyase; ATPase][Cyttoplasmic; Cytoskeletal] Skeletal muscle myosin heavy chain, member of a family of motor proteins that provide the force for muscle contraction, expressed only during embryogenesis Karsch-Mizrachi, I. et al., Nucleic Acids Res. 17:6167-79 (1989).
				[Schizosaccharomyces pombe] hypothetical protein with coiled-coil region; similar to S. cerevisiae YML071C; potential leucine zipper
				[Caenorhabditis elegans] Protein with weak similarity to H. sapiens Hs.177410 (Human GAP SH3 binding protein mRNA, complete cds)
				Jiang, M. et al. Proc. Natl. Acad. Sci. U.S.A. 98:218-223 (2001)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	6024712CD1	371	S11 S15 S294 S358 T52 T206 T259 T276 T348 T361	N127 N181	Signal_cleavage: M1-A29	SPSCAN
					lactate/malate dehydrogenase: K61-K370	HMMER_PFAM
					Transmembrane domains: V62-L80, L148-Q165	TMAP
					L-lactate dehydrogenase BL00064: K61-K98, D121-P168, S176-H220, S221-L250, D262-S313, E325-L369	BLIMPS_BLOCKS
					Malate dehydrogenase protein	BLIMPS_BLOCKS
					BL00068: S129-F156, L172-G218, V239-D256	
					L-lactate dehydrogenase active site L_jdh.prf: F209-D256	PROFILES SCAN
					L-lactate dehydrogenase signature PR00086: K61-D85, E86-F110, I173-S193, I197-Q215, W227-W240	BLIMPS_PRINTS
					DEHYDROGENASE OXIDOREDUCTASE NAD	BLAST_PRODROM
					MALATE LLACTATE GLYCOLYSIS ACID	
					TRICARBOXYLIC CYCLE MULTIGENE	
					PD000350: K61-E364	
					L-LACTATE DEHYDROGENASE DM00253	BLAST_DOMO
					P04642 18-330: S60-K370	
					P07864 17-329: K61-L369	
					I62761 19-331: H59-K370	
					P33571 20-332: H59-L369	
					L-lactate dehydrogenase active site: L:229-S235	MOTIFS

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	72176922CD1	236	S128 S164 S165 S169 S174 S189 S197 T221 Y68 S37 S73 T46 Y94 T118 T170 T205	N44 N55 N64	Signal Peptide: M1-G16	HMMER
					Immunoglobulin domain: E42-Y98	HMMER_PFAM
					Transmembrane domain: T133-R158	TMAP
					N-terminus is non-cytosolic	
3	1392717CD1	107	S5 S44 S84	N82	signal_cleavage: M1-A36	SPSCAN
					Transmembrane domain: T21-V42	TMAP
					N-terminus is non-cytosolic	
4	2701254CD1	124	S76		signal_cleavage: M1-A56	SPSCAN
					Signal Peptide: M1-P27, M1-A31	HMMER
5	71774318CD1	144	S81		signal_cleavage: M1-G29	SPSCAN
					Signal Peptide: M1-P34	HMMER
6	71802522CD1	202	S11 S19 S53 S82 S160 T2 T109 T151		signal_cleavage: M1-G52	SPSCAN
					Signal Peptide: L30-A58	HMMER
					Transmembrane domain: P33-S53, V119-W142	TMAP
					N-terminus is non-cytosolic	
7	6425956CD1	207	S82 S84 S115 S148 S166 T134		signal_cleavage: M1-G50	SPSCAN
					Signal Peptide: M1-P46, R21-G50	HMMER
					Transmembrane domain: H27-G52	TMAP
					N-terminus is non-cytosolic	
8	7494288CD1	291	S96 S143 T6 T86 Y214	N112 N141 N167	signal_cleavage: M26-S75	SPSCAN



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	(cont.)				Signal Peptide: P54-T79, M44-S80, M61-T79	HMMER
					Leucine Rich Repeat: Q206-P229, D109-M132, E133-K156, R157-H180, G181-M205	HMMER_PFAM
					Leucine rich repeat C-terminal domain: N239-K283	HMMER_PFAM
					Leucine rich repeat N-terminal domain: S80-P107	HMMER_PFAM
					Transmembrane domain: P56-H74	TMAP
					N-terminus is non-cytosolic	
					Leucine-rich repeat signature PR00019:A155-F168, L158-V171	BLIMPS_PRINTS
					Leucine zipper pattern: L137-L158	MOTIFS
9	7474330CD1	356	S205 S221 S226 S263 S268 S287 S289 S311 T137 T178 T237 T267		Connexin: M1-S205	HMMER_PFAM
					Signal Peptide: M1-Q32	HMMER
					Transmembrane domain: S4-R28, L61-L89, S126-L154, E179-V203	TMAP
					Connexins proteins BL00407: P57-H84, P123-G152, C161-S205, A26-S56	BLIMPS_BLOCKS
					Connexins signatures connexins_1.pr: M20-V71	PROFILESKAN
					Connexins signatures connexins_2.pr: A141-L197	PROFILESKAN
					Connexin signature PR00206: P7-Y31, F38-H60, F63-L83, F125-F151, C161-S181, L182-S205	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	(cont.)				GAP JUNCTION CONNEXIN PROTEIN TRANSMEMBRANE ALPHA1 CX43 ALPHA8 ALPHA5 BETA1 PD001135: S4-R85, Y129-L202	BLAST_PRODOM
					CONNEXINS DM00590 P35212 1-278: S4-S221	BLAST_DOMO
					CONNEXINS DM00590 P18860 1-278: S4-L202	BLAST_DOMO
					CONNEXINS DM00590 P28228 1-304: S4-L89	BLAST_DOMO
					CONNEXINS DM00590 P41987 1-277: S4-S226	BLAST_DOMO
					Connexins signature 1: C40-D53	MOTIFS
					Connexins signature 2: C161-P177	MOTIFS
10	5911370CD1	82	S21 S51		signal_cleavage: M1-G25	SPSCAN
					Signal Peptide: M1-D24, M1-G25, M1-T27, M1-M31	HMMER
					Gamma-thionins family: R36-C82	HMMER_PFAM
					Transmembrane domain: 14-Y23	TMAP
					N-terminus is non-cytosolic	
					Gamma-thionins family proteins BL00940: R36-C59, C71-C82	BLIMPS_BLOCKS
					GAMMA-THIONINS FAMILY Y DM00833 P21923 1-46: R36-C82	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): A35-S42	MOTIFS
					Gamma-thionins family signature: R36-C59	MOTIFS
11	7647134CD1	529	S32 S75 S175 S294 S340 S344 S370 S509 T38 T43 T190 T324 T409 T427 Y307 Y371	N224 N416	Signal Peptide: M1-S27	HMMER
					Signal Peptide: M1-C28	HMMER

Table 3

SEQ ID NO: Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11 (cont.)				signal_cleavage: M1-A26	SPSCAN
				'Paired box' domain: G114-R238	HMMER_PFAM
				Paired box' domain proteins BL00034: G114-S164, G168-N204, F208-R238, S269-P279	BLIMPS_BLOCKS
				'Paired box' domain signature: G128-S184	PROFILES SCAN
				Paired box signature PR00027: V118-D133, R136-R154, L156-T173, G174-P191	BLIMPS_PRINTS
				PROTEIN PAIRED BOX NUCLEAR DNA-BINDING DEVELOPMENTAL HOMEOBOX TRANSCRIPTION REGULATION PAX6 PD000643: G114-R238	BLAST_PRODROM
				PROTEIN PAIRED BOX DNA-BINDING DEVELOPMENTAL NUCLEAR TRANSCRIPTION REGULATION DIFFERENTIATION ALTERNATIVE PD002426: G414-P493	BLAST_PRODROM
				PAIRED BOX PROTEIN PAIRED BOX DNA-BINDING DEVELOPMENTAL PROTEIN NUCLEAR PROTEIN PD072729: P334-N410	BLAST_PRODROM
				PROTEIN PAIRED BOX DNA-BINDING DEVELOPMENTAL NUCLEAR TRANSCRIPTION REGULATION DIFFERENTIATION ALTERNATIVE PD004047: P334-N410	BLAST_PRODROM
				PAIRED BOX DM00579 Q02962 13-126: M111-D225	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					PAIRED BOX DM00579 S36156 I2-125: A110-D225	BLAST_DOMO
					PAIRED BOX DM00579 Q02548 I3-126: G114-D225	BLAST_DOMO
					PAIRED BOX DM00579 Q02650 I3-126: G114-D225	BLAST_DOMO
					'Paired box' domain signature: R148-S164	MOTIFS
12	1631327CD1	453	S92 S107 S111 S120 S149 S297 S426 T28 T174 T345 T414	N35	Signal Peptide: M1-V20	HMMER
					Signal Peptide: M1-S21, M1-L23, M1-W24, M1-A19	HMMER
					Transmembrane domain: W354-L382 N-terminus is cytosolic	TMAP
13	044232CD1	271	S17 S132 S160 S229 S230 S253 T191 T260 Y244	N85 N218	Signal Peptide: L54-C72	HMMER
					Signal Peptide: L54-Y74, L53-S73, V51-C72,	HMMER
					Signal_cleavage: M7-C72	SPSCAN
					Transmembrane domain: F42-V70, L90-Y113, G165-K193 N-terminus is non-cytosolic	TMAP
14	560293CD1	203	S21 S73 T6 T144 Y83		Signal Peptide: P19-A46	HMMER
					Signal_cleavage: M1-A56	SPSCAN

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	2025618CD1	529	S104 S116 S195 S206 S216 S307 S370 S464 T65 T119 T168 T187 T208 T228 T270 T283 T313 T499 T520	N63 N480	Signal Peptide: M1-V20, M1-A19, M1-A23, M1-G24, M1-T27	HMMER
					Transmembrane domain: L4-T29	TMAP
					Signal_cleavage: M1-G18	SPSCAN
					PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATP-BINDING FILAMENT HEPTAD PD000002: M272-E478	BLAST_PRODOM
					PROTEIN REPEAT TROPOMYOSIN COILED COIL ALTERNATIVE SPLICING SIGNAL PRECURSOR CHAIN PD000023: E273-N479	BLAST_PRODOM
					TROPOMYOSIN DM00077 P53933 S580-755: L300-Q467	BLAST_DOMO
					CALDESMON DM06224 P12957 1-755: Q95-N479	BLAST_DOMO
					TRICHOHYALIN DM03839 P37709 632-1103: K59-K476	BLAST_DOMO
16	3342443CD1	305	S143 S238 S286 T22 T51 T67 T150 T249	N78 N82	signal_cleavage: M1-G20	SPSCAN
					Signal Peptide: M1-G20, M1-R19, M1-Q21, M1-Q24, M1-E26	HMMER
					u-PAR/Ly-6 domain: T124-P140	HMMER_PFAM
					Ly-6 / u-PAR domain proteins BL00983: S59-C68, E122-N137	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	2267957CD1	493	S194 S243 S256 S477 T124 T188 T347 T469	N72 N264 N315 N349 N360	Signal Peptide: P7-A27, M1-R26, M1-A30, L10-R29, M1-R29  signal_cleavage: M1-A27 Leucine Rich Repeat: Q159-A182, V135-A158, Y62-R84, N111-Q134, K186-P209, Q87-P110 Leucine rich repeat C-terminal domain: N221-G271	HMMER  SPSCAN HMMER_PFAM HMMER_PFAM
					Immunoglobulin domain: G283-A343 Transmembrane domain: G9-A30 N371-W399 Leucine-rich repeat signature PF00019: L112-L125	HMMER_PFAM TMAP BLIMPS_PFAM
					Leucine zipper pattern L45-L66 L52-L73 L115-L136 L356-L377	MOTIFS
18	7480277CD1	869	S44 S78 S80 S121 S128 S181 S198 S257 S276 S344 S431 S449 S479 S668 S783 T49 T92 T187 T252 T311 T316 T463 T540 T648 T750 T761 T863	N42 N90 N131 N232 N455 N587 N666	Signal Peptide: M1-R17, M1- Y22	HMMER
					Signal_cleavage: M1-G18 MAM domain: C593-R758 Immunoglobulin domain: G53-A110, G353-V417, C601-S675, G150-A217, G256-T316 MAM domain proteins BL00740: C601-W613, L741-T761	SPSCAN HMMER_PFAM HMMER_PFAM BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18 (cont.)					MAM domain signature PR00020: K599-N617, Y672-K683, V720-G734, G739-K752	BLIMPS_PRINTS
					PRECURSOR GLYCOPROTEIN SIGNAL TRANSMEMBRANE HYDROLASE PROTEIN REPEAT RECEPTOR PHOSPHATASE NEUROPILIN PD001482: D590-C756	BLAST_PRODROM
					MAM DM01344 P28824 595-796: L552-D747	BLAST_DOMO
					PROTEIN-TYROSINE-PHOSPHATASE, RECEPTOR TYPE MU DM07136 P35822 1-187: P577-V749	BLAST_DOMO
					MAM DM01344 P98072 352-509: N587-D748	BLAST_DOMO
					MAM DM01344 A55620 618-796: T592-G742	BLAST_DOMO
19	3450647CDI	174	S42		Signal Peptide: M1-C18, M1-S21	HMIMER
					Signal Cleavage: M1-V19	SPSCAN
					Transmembrane domain: T84-Y108 H144-S163	TMAP
					N-terminus is non-cytosolic	
					Maspin Signature: S61-G79	BLIMPS_PRINTS
20	2053428CDI	561	S21 S83 S151 S164 S196 S216 S321 S515 T124 T266 T335 Y76 Y215 Y424	N337	Signal Peptide: M28-L45	HMIMER
					Signal Cleavage: M1-G46	SPSCAN
					Transmembrane domain: M454-R477	TMAP
					N-terminus is non-cytosolic	
					R02D3.2 PROTEIN PD147543: E79-G555	BLAST_PRODROM
					P1008 PROTEIN PD138971: D72-D322	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7503614CD1	219	S21 S73 T6 T171 Y83		signal_cleavage: M1-A56	SPSCAN
					Cytosolic domain: M215-D219 Transmembrane domain: P192-L214 Non-cytosolic domain: M1-G191	TMHMMER
					Ribosomal protein P2 signature PR00456: R24-S35, S35-A49	BLIMPS_PRINTS
22	7503456CD1	497	S104 S116 S195 S206 S216 S307 S370 T65 T119 T168 T187 T208 T228 T270 T283 T313 T467 T488	N63 N448	signal_cleavage: M1-G18	SPSCAN
					Signal Peptide: M1-G18, M1-A21, M1-G22, M1-A25, M1-G28	HMMEER
					Cytosolic domain: M1-T6 Transmembrane domain: V7-T29 Non-cytosolic domain: G30-V497	TMHMMER
					PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATP-BINDING FILAMENT HEPTAD PD000002: M272-K469, K260-K469	BLAST_PRODOR
					PROTEIN REPEAT TROPOMYOSIN COILED COIL ALTERNATIVE SPLICING SIGNAL PRECURSOR CHAIN PD000023: Q284-E446, K266-K469	BLAST_PRODOR
					TROPOMYOSIN DM00077 P53935 580-755: L300-L456	BLAST_DOMO



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23	7503459CD1	310	S21 S83 S151 S164 S196 S216 S296 S307 T124 T266 Y76 Y215		Signal Peptide: M28-A44, M28-G46	HMMER
					G-protein coupled receptors signature: S163-I179	MOTIFS
					Leucine zipper pattern: L88-L109	MOTIFS

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
24/6024712CB1/ 1197	1-455, 1-740, 82-211, 208-1197
25/72176922CB1/ 1001	1-148, 1-222, 1-228, 1-238, 1-266, 1-323, 1-376, 1-380, 1-431, 1-440, 1-445, 1-455, 1-466, 1-491, 1-499, 1-548, 1-554, 1-562, 1-565, 1-567, 1-573, 1-590, 1-608, 1-613, 1-630, 1-653, 1-664, 1-677, 1-687, 1-689, 1-711, 1-722, 5-453, 5-545, 11-233, 25-492, 38-560, 38-725, 42-208, 56-303, 149-427, 169-461, 202-427, 251-998, 285-761, 335-541, 335-977, 340-979, 359-588, 360-978, 370-622, 396-1001, 404-830, 419-973, 438-973, 448-984, 451-1000, 457-891, 466-540, 503-961, 506-1000, 535-973, 549-999, 557-973, 563-948, 570-989, 576-983, 580-998, 583-973, 612-1000, 621-1000, 633-976, 660-984, 661-1001, 671-984, 674-984, 713-973, 740-936, 740-984, 740-991, 794-984, 823-984, 852-971
26/1392717CB1/1174	1-267, 1-271, 1-753, 7-223, 116-269, 116-357, 116-387, 116-409, 116-412, 116-616, 120-381, 120-503, 125-412, 133-553, 135-374, 139-279, 150-330, 152-317, 154-371, 154-648, 155-281, 155-364, 156-679, 165-378, 166-420, 166-689, 168-295, 168-312, 168-319, 168-354, 171-311, 171-356, 171-359, 228-456, 266-880, 275-754, 308-563, 312-556, 354-603, 367-938, 386-694, 422-665, 423-675, 427-682, 469-1140, 474-696, 551-836, 574-889, 597-837, 597-847, 646-1107, 673-1116, 673-1147, 673-1156, 690-1113, 690-1174, 691-1165, 700-1159, 713-993, 720-973, 732-994, 736-1010, 744-1159, 745-1157, 776-1159, 778-929, 791-1157, 797-1139, 814-1156, 825-1157, 839-1097, 841-1157, 851-1157, 904-1155, 906-1174, 907-1155, 910-1157, 914-1145, 956-1157, 958-1156, 959-1120, 961-1166, 1008-1157, 1070-1174
27/2701254CB1/ 948	1-657, 229-662, 382-669, 382-948, 520-827
28/1774318CB1/ 2403	1-421, 100-766, 203-785, 224-893, 311-855, 372-892, 382-604, 415-1020, 424-1015, 435-1043, 444-994, 468-1069, 486-1178, 489-1160, 494-1171, 527-1069, 535-1125, 535-1159, 541-1062, 558-1070, 595-1123, 605-1227, 616-1200, 661-1197, 673-1250, 678-1341, 684-1269, 743-1480, 751-1282, 759-1336, 776-1327, 787-1435, 823-1352, 851-1370, 901-1603, 944-1635, 1006-1580, 1017-1570, 1043-1682, 1052-1629, 1065-1606, 1066-1684, 1073-1668, 1074-1742, 1083-1577, 1103-1758, 1107-1761, 1107-1764, 1119-1647, 1135-1749, 1173-1674, 1259-1881, 1263-1897, 1284-1942, 1289-1934, 1298-1960, 1308-2033, 1323-1951, 1368-2054, 1414-2001, 1433-2052, 1601-2252, 1603-2162, 1743-2248, 1758-2403, 1759-2299, 1774-2394, 1781-2318, 1823-2387

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
29/71802522CB1/ 2848	1-639, 442-925, 442-954, 442-1044, 442-1084, 461-1112, 466-991, 645-1285, 703-1371, 841-1478, 903-1474, 913-1526, 1087-1764, 1104-1739, 1203-1702, 1213-1862, 1290-1826, 1327-1849, 1362-1915, 1375-2011, 1398-2101, 1412-2009, 1419-2122, 1451-2028, 1466-2050, 1491-2083, 1510-2079, 1530-2081, 1552-2163, 1563-2028, 1565-2142, 1565-2202, 1583-2101, 1583-2127, 1612-2197, 1618-2266, 1638-2238, 1645-2305, 1687-2349, 1695-2362, 1708-2253, 1718-2083, 1762-2433, 1779-2246, 1844-1982, 1884-2543, 1885-2539, 1901-2521, 1913-2626, 1990-2551, 2104-2699, 2193-2848, 2240-2848, 2283-2848, 2297-2848, 2350-2848, 2366-2848, 2401-2805, 2542-2848
30/6425956CB1/ 3394	1-832, 240-835, 246-1096, 760-1000, 814-1457, 831-1108, 946-1308, 1179-1482, 1215-1905, 1254-1717, 1254-1807, 1254-1884, 1254-1923, 1267-1746, 1294-1626, 1397-1865, 1434-1865, 1692-1935, 1692-2219, 1692-2233, 1692-2246, 1692-2277, 1692-2325, 1692-2337, 1692-2376, 1692-2426, 1692-2434, 1692-2547, 1693-2052, 1696-2105, 1696-2274, 1696-2431, 1696-2453, 1697-2444, 1708-1823, 1802-2099, 1802-2396, 1824-2387, 1831-2073, 1850-2126, 1907-2469, 1912-2618, 1921-2324, 1929-2098, 1929-2404, 1931-2408, 1932-2560, 2001-2652, 2008-2277, 2008-2569, 2026-2477, 2037-2500, 2084-2728, 2096-2544, 2098-2275, 2101-2669, 2119-2679, 2144-2496, 2145-2679, 2158-2839, 2167-2714, 2178-2612, 2182-2438, 2205-2695, 2238-2772, 2250-2792, 2257-2849, 2301-2544, 2311-2851, 2313-2570, 2315-2616, 2324-2773, 2348-2607, 2370-2753, 2373-3053, 2375-2894, 2424-2838, 2440-2999, 2448-2782, 2448-2951, 2452-2792, 2455-2687, 2456-2968, 2464-2991, 2482-2703, 2488-2972, 2493-2860, 2497-2677, 2497-3004, 2498-2808, 2503-2907, 2503-3194, 2505-2961, 2518-2804, 2548-3133, 2549-2997, 2550-2603, 2551-2808, 2569-3010, 2573-3015, 2584-2849, 2586-2832, 2587-3014, 2596-3216, 2601-2809, 2618-3276, 2620-3122, 2651-3331, 2669-2832, 2677-3155, 2684-3100, 2687-3354, 2708-2948, 2710-3368, 2755-3027, 2765-3312, 2784-3060, 2790-3094, 2792-3394, 2814-3331, 2826-3336, 2835-3338, 2840-3364, 2852-3365, 2853-3387, 2865-3366, 2870-3394, 2871-3017, 2889-3368, 2893-3037, 2898-3394, 2925-3153, 2948-3394, 2958-3376, 2963-3366, 2969-3376, 2977-3376, 2983-3394, 2984-3366, 2987-3261, 3000-3317, 3011-3268, 3038-3369, 3038-3371, 3056-3170, 3057-3369, 3080-3391, 3109-3386, 3144-3368, 3146-3379, 3147-3394, 3174-3376, 3176-3394, 3248-3394
31/7494288CB1/ 1858	1-589, 1-692, 1-1858, 365-980, 421-646, 506-1195, 534-1176, 538-786, 539-1126, 951-1590, 958-1197, 961-1522, 963-1472, 966-1472, 1055-1331, 1440-1529
32/7474330CB1/ 1242	1-539, 179-1228, 1064-1242

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
33/5911370CB1/544	1-295, 1-313, 1-482, 1-524, 1-544, 9-491, 9-542, 9-544, 14-509, 38-488, 38-542, 38-544
34/7647134CB1/ 3471	1-3471, 510-586, 514-586, 520-958, 522-1159, 550-881, 587-711, 587-763, 587-825, 587-834, 587-870, 587-880, 587-914, 587-947, 587-949, 587-964, 587-978, 587-988, 587-1024, 587-1046, 589-828, 590-702, 592-661, 592-854, 618-685, 618-1114, 633-1136, 633-1156, 633-1159, 637-1159, 650-1159, 663-1159, 669-1159, 678-1159, 685-1159, 697-921, 707-1003, 707-1008, 720-916, 720-972, 723-1065, 723-1159, 727-972, 727-1159, 751-1159, 755-1159, 759-1159, 773-1048, 774-1159, 780-1159, 782-841, 786-1063, 812-1159, 826-1159, 827-1159, 843-1159, 851-1159, 859-1159, 860-958, 891-1159, 895-1159, 898-1159, 952-1159, 959-1140, 979-1159, 984-1087, 987-1159, 1007- 1159, 1010-1159, 1017-1159, 1029-1159, 1029-1166, 1032-1159, 1052-1159, 1069-1159, 1100-1159, 1118-1159, 1128-1159, 1129-1159, 1227-1250, 1227-1253, 1227-1255, 1227-1264, 1227-1266, 1227-1311, 1227-1342, 1227- 1363, 1227-1378, 1227-1380, 1227-1396, 1227-1406, 1227-1407, 1227-1414, 1227-1453, 1227-1490, 1227-1511, 1227-1530, 1227-1532, 1227-1533, 1227-1573, 1227-1609, 1227-1610, 1227-1631, 1227-1696, 1231-1533, 1240- 1533, 1242-1533, 1270-1407, 1272-1533, 1274-1533, 1280-1435, 1280-1533, 1301-1533, 1313-1533, 1318-1407, 1323-1533, 1324-1533, 1354-1533, 1360-1533, 1366-1533, 1385-1533, 1410-1911, 1452-1533, 2057- 2166, 2128-2315, 2263-2568, 2437-2682, 2437-2923, 2483-2724, 2488-3058, 2751-3040, 2810-3230, 3034-3286, 3079-3359, 3214-3360, 3303-3467
35/1631327CB1/ 1484	1-155, 1-569, 5-721, 6-235, 8-68, 15-284, 20-291, 20-296, 20-552, 21-245, 21-482, 27-647, 28-486, 29-300, 29-487, 30-498, 35-581, 37-640, 66-562, 69-279, 69-423, 69-716, 71-536, 80-724, 106-721, 132-326, 220-723, 224-723, 225- 724, 226-724, 227-704, 228-704, 228-723, 228-724, 229-684, 229-694, 230-684, 230-694, 231-687, 232-684, 232- 694, 233-675, 233-692, 233-694, 233-722, 233-724, 234-684, 234-694, 234-724, 236-723, 237-694, 238-724, 239- 724, 240-724, 241-724, 243-704, 246-704, 247-706, 249-719, 249-724, 251-684, 252-694, 254-704, 257-654, 258- 724, 280-706, 286-694, 286-724, 287-697, 287-724, 288-694, 288-724, 290-694, 291-694, 291-704, 292-690, 292- 694, 292-704, 292-706, 292-718, 297-724, 302-724, 304-704, 306-706, 308-694, 309-724, 310-724, 312-724, 314- 679, 318-676, 318-705, 321-694, 321-704, 321-723, 321-724, 321-869, 327-684, 333-694, 333-720, 336- 699, 336-706, 341-723, 342-694, 342-724, 345-694, 357-696, 359-694, 360-706, 361-682, 368-724, 372-505, 378- 724, 383-694, 389-538, 407-704, 411-723, 413-684, 413-694, 420-723, 427-538, 427-724, 429-538, 429-684, 430- 704, 431-720, 431-724, 434-694, 448-724, 449-704, 457-715, 458-704, 459-538, 462-694, 462-706, 467-538,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
35 (cont.)	478-538, 478-704, 480-712, 491-724, 492-720, 498-684, 514-694, 518-538, 528-1065, 538-694, 538-695, 538-704, 538-719, 538-723, 538-724, 539-724, 560-1237, 582-704, 643-1205, 680-1219, 829-1323, 837-1298, 849-1072, 866-1415, 901-1323, 913-1322, 932-1314, 933-1289, 940-1313, 940-1327, 1000-1484, 1003-1484, 1005-1322, 1032-1282, 1033-1298, 1072-1209
36/044232CB1/ 1773	1-350, 9-185, 28-288, 28-568, 29-269, 29-483, 29-584, 29-617, 29-631, 29-648, 29-668, 30-269, 30-463, 30-652, 31-650, 32-377, 33-390, 34-186, 43-557, 51-166, 51-283, 56-266, 61-294, 62-350, 63-385, 64-244, 69-672, 174-421, 202-701, 203-806, 203-809, 205-782, 225-797, 294-514, 531-730, 541-817, 587-1111, 588-874, 612-742, 626-909, 652-1305, 782-1225, 789-1264, 789-1387, 793-1068, 797-1365, 806-1416, 815-1441, 842-1092, 843-1137, 910-1441, 959-1244, 1049-1416, 1100-1361, 1307-1773, 1609-1686, 1699-1744
37/560293CB1/ 2016	1-504, 24-144, 38-327, 38-570, 42-352, 42-430, 42-524, 42-581, 42-615, 44-424, 62-389, 69-424, 76-646, 91-696, 120-169, 136-703, 146-440, 149-328, 172-595, 184-703, 191-685, 205-703, 208-671, 210-671, 211-667, 211-671, 213-804, 215-679, 217-679, 221-679, 222-679, 222-703, 223-679, 226-683, 230-679, 237-685, 244-671, 245-703, 247-632, 248-606, 250-641, 260-679, 262-685, 271-679, 274-634, 280-671, 280-679, 281-679, 282-671, 283-679, 286-700, 289-703, 295-671, 299-679, 300-703, 302-645, 302-703, 306-703, 312-703, 319-679, 323-701, 326-599, 327-635, 328-663, 338-675, 354-703, 364-671, 375-703, 378-703, 413-703, 421-703, 429-703, 434-703, 460-703, 476-703, 495-703, 515-703, 527-679, 540-1164, 583-647, 583-685, 587-663, 699-804, 728-931, 729-888, 729-989, 729-1020, 729-1080, 729-1149, 729-1153, 729-1159, 729-1197, 729-1200, 729-1203, 729-1206, 729-1216, 729-1224, 729-1226, 729-1283, 729-1287, 729-1355, 730-1257, 731-1072, 732-872, 732-888, 734-888, 734-1152, 734-1182, 737-1098, 737-1195, 738-1263, 741-1235, 741-1322, 747-1202, 747-1210, 748-984, 748-1081, 749-988, 750-1122, 756-1122, 757-1207, 759-1167, 760-955, 760-1237, 764-950, 766-1054, 776-976, 777-1204, 790-976,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
37 (cont.)	790-1031, 791-1291, 792-1413, 796-1322, 818-1110, 818-1150, 818-1159, 818-1195, 818-1196, 818-1208, 822-1074, 822-1508, 848-1059, 848-1091, 864-1203, 899-1217, 901-1153, 909-1513, 936-1516, 974-1175, 977-1239, 978-1189, 989-1249, 990-1231, 1010-1278, 1074-1347, 1088-1391, 1097-1239, 1121-1292, 1121-1367, 1135-1401, 1136-1305, 1152-1367, 1152-1408, 1154-1380, 1171-1349, 1171-1381, 1171-1394, 1179-1395, 1194-1451, 1200-1505, 1200-1540, 1202-1449, 1209-1489, 1254-1546, 1276-1425, 1306-1542, 1313-1523, 1314-1555, 1314-1852, 1343-1696, 1344-1582, 1345-1521, 1349-1570, 1360-1596, 1375-1569, 1392-1629, 1402-1646, 1402-1877, 1462-1691, 1473-1699, 1473-1789, 1473-1814, 1504-1720, 1517-1916, 1548-1805, 1554-1706, 1567-1809, 1573-1852, 1617-1851, 1767-2016
38/2025618CB1/ 2520	1-300, 1-508, 1-550, 28-449, 35-615, 216-771, 223-347, 312-395, 312-885, 349-596, 360-784, 365-899, 404-690, 453-1028, 475-744, 516-764, 534-803, 586-851, 658-1203, 696-1026, 810-1122, 967-1200, 967-1432, 1025-1197, 1026-1258, 1027-1440, 1031-1333, 1131-1359, 1136-1384, 1168-1562, 1168-1569, 1179-1609, 1288-1720, 1345-1914, 1353-1639, 1353-1649, 1399-1656, 1485-1778, 1490-1770, 1521-2204, 1535-1717, 1555-2164, 1590-2148, 1608-2203, 1614-1898, 1680-1907, 1688-1976, 1781-2374, 1904-2214, 1959-2383, 2002-2384, 2014-2219, 2056-2084, 2093-2321, 2102-2383, 2125-2219, 2248-2520
39/3342443CB1/ 1036	1-70, 1-207, 1-266, 20-269, 20-557, 24-256, 24-466, 69-138, 122-373, 400-1036, 469-953
40/2267957CB1/ 1621	1-1621, 120-1018, 120-1128, 128-756, 538-1235, 559-790, 559-1003, 559-1080, 1191-1431, 1191-1601
41/7480277CB1/ 3562	1-186, 1-946, 187-946, 630-1161, 819-1267, 852-1266, 1018-1334, 1207-1664, 1334-3562
42/3450647CB1/ 899	1-474, 1-502, 50-293, 50-433, 50-518, 50-541, 50-648, 50-778, 50-805, 50-809, 52-420, 52-635, 52-692, 52-727, 52-809, 52-885, 52-899, 55-832, 91-899, 253-899, 282-898, 303-899, 328-899, 359-899, 395-899, 423-500, 456-899, 482-899, 504-899, 568-899, 570-899

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
43/2053428CB1/ 2330	1-880, 532-699, 532-793, 642-899, 645-886, 645-889, 650-833, 650-871, 652-883, 652-903, 652-917, 653-858, 653-871, 653-909, 653-917, 653-924, 653-940, 653-994, 653-1059, 653-1239, 654-815, 654-948, 655-1027, 660-1268, 671-1351, 672-1158, 868-1511, 907-1630, 914-1498, 940-1245, 1047-1307, 1080-1362, 1097-1282, 1097-1452, 1097-1480, 1207-1805, 1233-1485, 1324-1928, 1355-1893, 1355-1903, 1366-2019, 1381-1637, 1381-1679, 1401-1902, 1417-1644, 1417-1760, 1446-1879, 1450-1677, 1474-1774, 1508-1731, 1590-1844, 1590-2105, 1603-1884, 1606-2148, 1607-1873, 1619-1837, 1619-1850, 1629-2245, 1645-1899, 1698-1977, 1700-1927, 1728-1928, 1765-2330, 1800-2054, 1809-2296, 1822-2210, 1881-2297, 1896-2306, 1899-2254, 1916-2306, 1918-2301, 1928-2305, 1945-2306, 1971-2306, 2019-2230, 2077-2301
44/7503614CB1/ 1755	1-202, 1-207, 1-310, 1-478, 1-1755, 12-301, 13-299, 13-544, 14-251, 16-165, 16-289, 16-326, 16-404, 16-555, 16-589, 51-620, 65-670, 110-669, 153-780, 153-797, 153-822, 153-828, 153-831, 153-840, 153-842, 153-879, 153-950, 158-659, 158-665, 161-663, 165-659, 178-677, 182-645, 184-645, 185-641, 185-645, 186-677, 189-653, 191-653, 195-653, 196-653, 196-677, 197-653, 200-657, 202-669, 204-653, 206-632, 211-659, 218-645, 219-667, 221-606, 222-580, 224-615, 234-653, 236-659, 245-653, 248-608, 254-645, 254-653, 255-653, 256-645, 257-653, 260-669, 263-677, 269-645, 273-653, 275-637, 276-619, 276-669, 280-669, 280-677, 286-677, 293-653, 301-609, 302-637, 312-649, 318-677, 338-645, 352-669, 354-543, 354-665, 370-600, 379-564, 380-591, 387-677, 395-677, 403-677, 408-677, 424-602, 428-600, 434-677, 456-669, 469-677, 489-677, 501-653, 703-950, 709-862, 712-862, 720-950, 734-929, 738-924, 750-1001, 764-950, 792-885, 984-1036, 1099-1354, 1141-1385, 1141-1616, 1282-1618, 1306-1548, 1362-1618, 1418-1618, 1466-1655, 1506-1755, 1531-1657
45/7503456CB1/ 2427	1-511, 3-2427, 1421-1618, 1421-1619, 1436-2111, 1439-1664, 1439-1703, 1439-1707, 1462-2071, 1490-2086, 1497-2055, 1515-2110, 1521-1805, 1548-2396, 1550-2396, 1554-2396, 1557-2389, 1568-2395, 1587-1814, 1595-1883, 1601-2396, 1654-2396, 1657-2126, 1667-2112, 1679-1795, 1680-2126, 1681-2386, 1685-2129, 1697-2118, 1706-2122, 1811-2121, 1866-2290, 1909-2291, 1921-2126, 2000-2228, 2009-2290, 2155-2427
46/7503459CB1/ 1685	1-1661, 7-566, 322-1042, 329-585, 331-1040, 331-1073, 332-575, 336-557, 338-569, 339-610, 339-626, 339-680, 339-925, 340-1042, 341-501, 341-634, 341-745, 347-954, 357-1037, 358-844, 626-931, 734-993, 766-1048, 783-968, 848-1087, 859-1080, 1024-1284, 1030-1285, 1109-1283, 1120-1685, 1142-1615, 1155-1409, 1164-1651, 1180-1565, 1212-1660, 1216-1337, 1236-1652, 1251-1661, 1254-1609, 1273-1656, 1283-1660, 1300-1661, 1326-1661, 1374-1585, 1432-1656

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
24	6024712CB1	TESTNOT11
25	72176922CB1	EOSIHET02
26	1392717CB1	THYRNOT03
27	2701254CB1	OVARTUT10
29	71802522CB1	UTRSNOR01
30	6425956CB1	LUNGNON07
31	7494288CB1	BRAIFER05
33	5911370CB1	BRAIFEN03
34	7647134CB1	KIDCTME01
35	1631327CB1	SINTNOR01
36	044232CB1	OVARDIR01
37	560293CB1	LUNGFET03
38	2025618CB1	LUNGNON03
39	3342443CB1	SPLNNOT09
40	2267957CB1	UTRSNOT02
41	7480277CB1	ADRETUE04
42	3450647CB1	UTRSNON03
43	2053428CB1	PROSTUS20
44	7503614CB1	COLHTUS02
45	7503456CB1	BRAINOT09
46	7503459CB1	293TF1T01



Table 6

Library	Vector	Library Description
293TF1T01	pINCY	Library was constructed using RNA isolated from a transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue. The cells were transformed with adenovirus 5 DNA.
ADRETUE04	PCDNA2.1	This 5 prime biased random primed library was constructed using RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma. Patient history included benign hypertension, depressive disorder, chronic sinusitis, idiopathic proctocolitis, a cataract, and urinary tract infection. Previous surgeries included a vaginal hysterectomy. Patient medications included Procardia (one dose only) and Prozac for 5 years. Family history included secondary Parkinsonism in the father; cerebrovascular disease, secondary Parkinsonism and anxiety state in the mother; and benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and brain cancer in the sibling(s).
BRAIFEN03	pINCY	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAINOT09	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus, who died at 23 weeks gestation.
COLHTUS02	pINCY	This subtracted colon tumor tissue library was constructed using 4.24 million clones from a colon tumor library and was subjected to two rounds of subtraction hybridization with 4.04 million clones from an ascending/transverse colon tissue library. The starting library for subtraction was constructed using RNA isolated from colon tumor tissue removed from the hepatic flexure of a 55-year-old Caucasian male during right hemicolectomy, incidental appendectomy, and permanent colostomy. Pathology indicated invasive grade 3 adenocarcinoma that formed a circumferential mass in the ascending colon, located 10.5 cm from the distal resection margin. The tumor infiltrated through the muscularis propria into the pericolonic adipose tissue to within 0.4 cm of the radial fat margin. Patient history included benign hypertension, anxiety, abnormal blood chemistry, blepharitis, heart block, osteoporosis, and hyperplasia of prostate. Family history included prostate cancer, acute myocardial infarction, stroke, and atherosclerotic coronary artery disease. The hybridization probe for subtraction was derived from a similarly constructed library using RNA isolated from non-tumorous ascending and transverse colon tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al., Genome Research 6 (1996):791.

Table 6

Library	Vector	Library Description
EOSIHE02	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hyper eosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
KIDCTME01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology indicated the margins of resection were free of involvement. Pathology for the matched tumor tissue indicated grade 3 renal cell carcinoma, clear cell type, forming a variegated multicystic mass situated within the mid-portion of the kidney. The tumor invaded deeply into but not through the renal capsule.
LUNGFET03	pINCY	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
LUNGNON03	PSPORT1	This normalized library was constructed from 2.36 million independent clones from a lung tissue library. RNA was made from lung tissue removed from the left lobe a 58-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Patient also received radiation therapy to the retroperitoneum. Family history included prostate cancer, breast cancer, and acute leukemia. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228; Swaroop et al., NAR (1991) 19:1954; and Bonaldo et al., Genome Research (1996) 6:791.
LUNGNON07	pINCY	This normalized lung tissue library was constructed from 5.1 million independent clones from a lung tissue library. Starting RNA was made from RNA isolated from lung tissue. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
OVARDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from right ovary tissue removed from a 45-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, vaginal suspension and fixation, and incidental appendectomy. Pathology indicated stromal hyperthecosis of the right and left ovaries. Pathology for the matched tumor tissue indicated a dermoid cyst (benign cystic teratoma) in the left ovary. Multiple (3) intramural leiomyomata were identified. The cervix showed squamous metaplasia. Patient history included metrorrhagia, female stress incontinence, alopecia, depressive disorder, pneumonia, normal delivery, and deficiency anemia. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and primary tuberculous complex.

Table 6

Library	Vector	Library Description
OVARTUT10	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 58-year-old Caucasian female during a total abdominal hysterectomy, removal of a solitary ovary, and repair of inguinal hernia. Pathology indicated a metastatic grade 3 adenocarcinoma of colonic origin, forming a partially cystic and necrotic tumor mass in the left ovary, and an adenocarcinoma of colonic origin, forming a nodule in the left mesovarium. A single intramural leiomyoma was identified in the myometrium. The cervix showed mild chronic cystic cervicitis. Patient history included benign hypertension, follicular cyst of the ovary, colon cancer, benign colon neoplasm, and osteoarthritis. Family history included emphysema, myocardial infarction, atherosclerotic coronary artery disease, benign hypertension, and hyperlipidemia.
PROSTUS20	pINCY	This subtracted prostate tumor tissue library was constructed using 2.36 million clones from the PROSTUT13 library and was subjected to two rounds of subtraction hybridization with 1.56 million clones from FIBNOT01. The starting library for subtraction was constructed using RNA isolated from prostate tumor tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) involving the prostate peripherally with invasion of the capsule. Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen. Patient history included diverticulitis of colon, asbestosis, and thrombophlebitis. Family history included benign hypertension, multiple myeloma, hyperlipidemia, and rheumatoid arthritis. The hybridization probe for subtraction was derived from a similarly constructed library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al. Genome Research (1996) 6:791.
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
SPLNNOT09	pINCY	Library was constructed using RNA isolated from diseased spleen tissue removed from a 22-year-old Caucasian male (Ashkenazi Jewish descent) during a total splenectomy. Pathology indicated Gaucher's disease with marked splenomegaly. The patient presented with thrombocytopenia and congenital anomaly of the spleen. Patient history included benign disorders and type 1 Gaucher's disease. Patient medications included Synthroid. Family history included benign hypertension, thyroid disease, and a complete thyroidectomy in the mother; thyroid disease in the sibling(s); and benign hypertension, myocardial infarction, cerebrovascular disease, arteriosclerotic cardiovascular disease, and prostate cancer in the grandparent(s).
TESTNOT11	pINCY	Library was constructed using RNA isolated from testicular tissue removed from a 16-year-old Caucasian male who died from hanging. Patient history included drug use (tobacco, marijuana, and cocaine use), and medications included Lithium, Ritalin, and Paxil.

Table 6

Library	Vector	Library Description
THYRNOT03	pINCY	Library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.
UTRSNON03	pINCY	This normalized library was constructed from 6.4M independent clones from the UTRSNOT12 library. RNA was isolated from uterine myometrial tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Pathology for the associated tumor tissue indicated uterine leiomyoma. Patient history included ventral hernia and a benign ovarian neoplasm. The normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9228).
UTRSNOR01	pINCY	Library was constructed using RNA isolated from uterine endometrium tissue removed from a 29-year-old Caucasian female during a vaginal hysterectomy and cystocele repair. Pathology indicated the endometrium was secretory, and the cervix showed mild chronic cervicitis with focal squamous metaplasia. Pathology for the associated tumor tissue indicated intramural uterine leiomyoma. Patient history included hypothyroidism, pelvic floor relaxation, and paraplegia. Family history included benign hypertension, type II diabetes, and hyperlipidemia.
UTRSNOT02	PSPORT1	Library was constructed using RNA isolated from uterine tissue removed from a 34-year-old Caucasian female during a vaginal hysterectomy. Patient history included mitral valve disorder. Family history included stomach cancer, congenital heart anomaly, irritable bowel syndrome, ulcerative colitis, colon cancer, cerebrovascular disease, type II diabetes, and depression.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value=1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART, or TIGRFAM hits: Probability value=1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

- a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23,
- b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2-23,
- c) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to the amino acid sequence of SEQ ID NO:1,
- d) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, and
- e) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant

polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

- b) recovering the polypeptide so expressed.

5 10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

10 12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:24-46,  
b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of  
15 SEQ ID NO:25-46,  
c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to the polynucleotide sequence of SEQ ID NO:24,  
d) a polynucleotide complementary to a polynucleotide of a),  
e) a polynucleotide complementary to a polynucleotide of b),  
20 f) a polynucleotide complementary to a polynucleotide of c), and  
g) an RNA equivalent of a)-f).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

25

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under  
30 conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and  
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

35



15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 5           a)       amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b)       detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

10           17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

15           19. A method for treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment the composition of claim 17.

20           20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a)       exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b)       detecting agonist activity in the sample.

25           21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional SECP, comprising administering to a patient in need of such treatment a composition of claim 21.

30           23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a)       exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 35           b)       detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional SECP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of SECP in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of SECP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of SECP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

5 36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- 10 b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

15 37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

20 39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-23, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- 25 c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of
- 30 SEQ ID NO:1-23.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-23.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first

oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

5           49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

          50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

10

          51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

          52. An array of claim 48, which is a microarray.

15

          53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

          54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to  
20 said solid substrate.

          55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains  
25 nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

          56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

30           57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

          58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

          59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

35

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

5 62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

10 65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

15 67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

20 70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

25 72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

30 75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

35 77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

5

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

10

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

15

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

20

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

25

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

30

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

35



90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
5 NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:37.

10 93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:39.

15 95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
20 NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:42.

25 98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:44.

30 100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID  
35 NO:46.

<110> INCYTE GENOMICS, INC.

YUE, Henry  
 YANG, Junming  
 ELLIOTT, Vicki S.  
 DUGGAN, Brendan M.  
 HONCHELL, Cynthia D.  
 LEE, Sally  
 THANGAVELU, Kavitha  
 GIETZEN, Kimberly J.  
 FORSYTHE, Ian J.  
 LU, Dyung Aina M.  
 GRIFFIN, Jennifer A.  
 GURURAJAN, Rajagopal  
 LAL, Preeti G.  
 BAUGHN, Mariah R.  
 XU, Yuming  
 TANG, Y. Tom  
 AZIMZAI, Yalda  
 AU-YOUNG, Janice  
 KALLICK, Deborah A.  
 WALIA, Narinder K.  
 MASON, Patricia M.  
 TRAN, Uyen K.

<120> SECRETED PROTEINS

<130> PI-0394 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/273,946; 60/276,873; 60/280,531; 60/280,596; 60/332,426; 60/334,229;  
 60/347,703

<151> 2001-03-06; 2001-03-16; 2001-03-30; 2001-03-30; 2001-11-16; 2001-11-28;  
 2002-01-11

<160> 46

<170> PERL Program

<210> 1

<211> 371

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 6024712CD1

<400> 1

Met	Ser	Trp	Thr	Val	Ser	Val	Val	Gln	Ala	Ser	Arg	Arg	Val	Ser
1				5				10					15	
Ser	Ala	Gly	Ala	Asn	Phe	Leu	Ser	Leu	Cys	Pro	Ser	Gln	Ala	Ala
				20				25					30	
Arg	Met	Pro	Leu	Lys	Gly	Ala	Trp	Leu	Phe	Thr	Pro	Val	Lys	Ser
				35				40					45	
Glu	Leu	Val	Glu	Arg	Phe	Thr	Ser	Glu	Glu	Pro	Ala	His	His	Ser
				50				55					60	
Lys	Val	Ser	Ile	Ile	Gly	Thr	Gly	Ser	Val	Gly	Met	Ala	Cys	Ala
				65				70					75	
Thr	Ser	Ile	Leu	Leu	Lys	Gly	Leu	Ser	Asp	Glu	Leu	Ala	Leu	Val
				80				85					90	
Asp	Leu	Asp	Glu	Gly	Lys	Leu	Lys	Gly	Glu	Thr	Met	Asp	Leu	Gln

	95		100		105
His Gly Ser Pro	Phe Met Lys Thr Pro	Asn Ile Val Cys Ser	Lys		
	110		115		120
Asp Tyr Leu Val	Thr Ala Asn Ser Ser	Leu Val Ile Ile Thr	Glu		
	125		130		135
Gly Ala Arg Gln	Glu Lys Gly Glu Thr	Arg Leu Asn Leu Val	Gln		
	140		145		150
Arg Asn Val Ala	Ile Phe Lys Leu Met	Ile Ser Gly Ile Val	Gln		
	155		160		165
Tyr Ser Pro Leu	Cys Lys Leu Ile Ile	Val Ser Asn Pro Val	Asp		
	170		175		180
Asn Leu Thr Tyr	Val Ala Trp Lys Leu	Ser Ala Phe Ser Lys	Asn		
	185		190		195
Arg Ile Ile Gly	Ser Gly Cys Asn Leu	Asp Thr Ala Arg Phe	Arg		
	200		205		210
Phe Leu Ile Gly	Gln Lys Leu Gly Ile	His Ser Glu Ser Cys	His		
	215		220		225
Gly Trp Ile Leu	Gly Glu His Gly Asp	Ser Ser Val Pro Val	Trp		
	230		235		240
Ser Gly Val Asn	Ile Ala Gly Val Pro	Leu Lys Asp Leu Asn	Ser		
	245		250		255
Asp Ile Gly Thr	Asp Lys Asp Pro Glu	Gln Trp Lys Asn Val	His		
	260		265		270
Lys Glu Val Thr	Ala Thr Ala Tyr Glu	Ile Ile Lys Met Lys	Gly		
	275		280		285
Tyr Thr Ser Trp	Ala Ile Gly Leu Ser	Val Ala Asp Leu Thr	Glu		
	290		295		300
Ser Ile Leu Lys	Asn Leu Arg Arg Ile	His Pro Val Ser Thr	Ile		
	305		310		315
Ile Lys Gly Leu	Tyr Gly Ile Asp Glu	Glu Val Phe Leu Ser	Ile		
	320		325		330
Pro Cys Ile Leu	Gly Glu Asn Gly Ile	Thr Asn Leu Ile Lys	Ile		
	335		340		345
Lys Leu Thr Pro	Glu Glu Glu Ala His	Leu Lys Lys Ser Ala	Lys		
	350		355		360
Thr Leu Trp Glu	Ile Gln Asn Lys Leu	Lys Leu			
	365		370		

&lt;210&gt; 2

&lt;211&gt; 236

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72176922CD1

&lt;400&gt; 2

Met Thr Ala Glu	Phe Leu Ser Leu Leu	Cys Leu Gly Leu Cys Leu		
1	5	10	15	
Gly Tyr Glu Asp	Glu Lys Lys Asn Glu	Lys Pro Pro Lys Pro Ser		
	20	25	30	
Leu His Ala Trp	Pro Ser Ser Val Val	Glu Ala Glu Ser Asn Val		
	35	40	45	
Thr Leu Lys Cys	Gln Ala His Ser Gln	Asn Val Thr Phe Val Leu		
	50	55	60	
Arg Lys Val Asn	Asp Ser Gly Tyr Lys	Gln Glu Gln Ser Ser Ala		
	65	70	75	
Glu Asn Glu Ala	Glu Phe Pro Phe Thr	Asp Leu Lys Pro Lys Asp		
	80	85	90	
Ala Gly Arg Tyr	Phe Cys Ala Tyr Lys	Thr Thr Ala Ser His Glu		
	95	100	105	
Trp Ser Glu Ser	Ser Glu His Leu Gln	Leu Val Val Thr Asp Lys		

	110		115		120
His Asp Glu Leu Glu Ala Pro Ser Met Lys Thr Asp Thr Arg Thr					
	125		130		135
Ile Phe Val Ala Ile Phe Ser Cys Ile Ser Ile Leu Leu Leu Phe					
	140		145		150
Leu Ser Val Phe Ile Ile Tyr Arg Cys Ser Gln His Gly Ser Ser					
	155		160		165
Ser Glu Glu Ser Thr Lys Arg Thr Ser His Ser Lys Leu Pro Glu					
	170		175		180
Gln Glu Ala Ala Glu Ala Asp Leu Ser Asn Met Glu Arg Val Ser					
	185		190		195
Leu Ser Thr Ala Asp Pro Gln Gly Val Thr Tyr Ala Glu Leu Ser					
	200		205		210
Thr Ser Ala Leu Ser Glu Ala Ala Ser Asp Thr Thr Gln Glu Pro					
	215		220		225
Pro Gly Ser His Glu Tyr Ala Ala Leu Lys Val					
	230		235		

<210> 3  
 <211> 107  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1392717CD1

<400> 3	
Met Lys Pro Ser Ser Pro Arg Glu Trp Gly Glu Gln Glu His Cys	
1 5 10 15	
Thr Ser Pro Gln Trp Thr Leu Trp Ser Leu Ser Ala Val Ala Phe	
20 25 30	
Gln Gly Trp Ala Leu Ala Arg Ala Pro Val Ala Val Ser Ser Phe	
35 40 45	
Ala Asp Pro Asp Gln Lys Ser Leu Gln Thr Asn Leu Leu Leu Glu	
50 55 60	
Leu Arg Gly Arg Trp His Asn Arg Arg Ser Asp Gly Cys Arg Met	
65 70 75	
Cys Trp Thr Tyr Ile Ala Asn Arg Ser Leu Val Glu Gly Asp Ile	
80 85 90	
Leu Thr Lys Cys Pro Asp Leu Glu Val Ala Phe Leu Thr Trp Leu	
95 100 105	
Leu Val	

<210> 4  
 <211> 124  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2701254CD1

<400> 4	
Met Leu Thr Gln Ser Gln Gln Val Leu Arg Gly Ile Leu Leu Phe	
1 5 10 15	
Leu Gln Asn Ile Leu Gln Val Ser Trp Gly Ser Pro Leu Ala Leu	
20 25 30	
Ala Ser Pro Pro Ser Pro Ser Leu Gln Pro Gly Asn Gly Leu Ala	
35 40 45	
Ser Ser Leu Leu Ala Leu Gln Pro Gly Leu Ala Gly Pro Trp Ala	
50 55 60	

Gly	Pro	Gln	Glu	Pro	Ser	Pro	Ala	Met	Cys	Phe	Pro	Lys	Lys	Arg
				65					70					75
Ser	Leu	Trp	Pro	Asn	Leu	Arg	Lys	Gln	Trp	Ala	Ser	Ile	His	Ile
				80					85					90
Asn	Asp	Pro	Arg	Gly	Thr	Leu	Cys	Pro	Arg	Cys	Thr	Gly	Cys	Asn
				95					100					105
Gln	Arg	Gly	Ser	Gly	Gly	Ser	Gly	Leu	Ile	Trp	Arg	Asp	Arg	Phe
				110					115					120
Tyr	His	His	Pro											

&lt;210&gt; 5

&lt;211&gt; 144

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71774318CD1

&lt;400&gt; 5

Met	Leu	Phe	Pro	Ala	Gly	Thr	Leu	Ser	Leu	Ser	Pro	Gln	Pro	Tyr
1				5					10					15
Arg	Thr	Pro	Val	Leu	Ala	Ser	Phe	Trp	Phe	Pro	Cys	Leu	Gly	His
				20					25					30
Pro	Val	His	Pro	Gln	Val	Gly	Leu	Cys	Leu	Ser	Gln	Gly	Gln	Ser
				35					40					45
Cys	Leu	Ser	Leu	Pro	Arg	Thr	Ala	Gln	His	Ala	Ser	Ala	Gln	Ala
				50					55					60
Ser	Gly	Pro	Cys	Pro	Arg	Gly	Ser	Gly	Pro	Arg	Val	Trp	His	Cys
				65					70					75
His	Ser	Glu	Ala	Trp	Ser	Trp	Lys	Lys	Gly	Pro	Ser	Trp	Gln	Pro
				80					85					90
Phe	Glu	Gln	Pro	Pro	Ser	Pro	Ser	His	Phe	Leu	Glu	Pro	Ser	Pro
				95					100					105
Leu	His	Thr	Leu	Asp	Ser	Trp	Tyr	Leu	Thr	Ala	Ala	Val	Leu	Gly
				110					115					120
Glu	Thr	Trp	Pro	Ala	Ala	Thr	Phe	Pro	Arg	Phe	Glu	Lys	Lys	Leu
				125					130					135
Phe	Val	Ser	Phe	Tyr	Ile	Leu	Lys	Leu						
				140										

&lt;210&gt; 6

&lt;211&gt; 202

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71802522CD1

&lt;400&gt; 6

Met	Thr	Pro	Arg	Leu	Phe	Leu	Phe	Ser	Lys	Ser	Pro	Arg	Tyr	Arg
1				5					10					15
Ala	Gly	His	Ser	Gly	Arg	Gly	Ala	Gln	His	Leu	Leu	Pro	Asp	Leu
				20					25					30
Gly	Leu	Pro	Trp	Leu	Ser	Leu	Pro	Ala	Pro	Leu	Cys	Phe	Phe	Phe
				35					40					45
Ala	Ser	Pro	Leu	Ser	Leu	Gly	Ser	Pro	Lys	Ile	Ser	Ala	Thr	Ala
				50					55					60
Pro	Thr	Phe	His	Pro	Ala	Gln	Ala	Thr	Trp	Gln	Cys	Cys	Leu	Phe
				65					70					75
Gly	Leu	Gln	Met	Leu	Cys	Ser	Pro	Lys	Pro	Ser	Leu	Thr	Met	Thr

	80		85		90
Phe Ile Leu Ala	Pro Glu Cys Ser	Pro Gln Arg Ala	Lys Leu Gly		
	95		100		105
Ala Lys His Thr	Gln Lys Leu Gly	Gly Gly Lys Gly	Ala Val Lys		
	110		115		120
Trp Arg Trp Leu	Gly Arg Arg Ala	Leu Thr Ile Leu	Ile Ala Lys		
	125		130		135
Val Thr Leu Gly	Leu Trp Trp Gly	Gly Ala Glu Ala	His Ser Leu		
	140		145		150
Thr Ser Trp Asp	Leu Pro Glu Pro	Ala Ser Pro Thr	Glu Leu Gly		
	155		160		165
Gln Leu Leu Gln	Ser Val Glu Leu	Ala Phe Pro Leu	Phe Gly Glu		
	170		175		180
Gly Phe Gly Ile	Trp Gly Phe Arg	Ser Pro Gly Lys	Val Arg Val		
	185		190		195
Leu Cys Thr Gln	Ala Pro Ala				
	200				

&lt;210&gt; 7

&lt;211&gt; 207

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6425956CD1

&lt;400&gt; 7

Met Gly Lys Gly	Gly Leu Ala His	Gly Ala Gly	Leu Leu Val	Leu	
1	5	10		15	
Pro Glu His Gly	Gly Arg Gly Ala	Pro Ala Leu	His Gln Ala	Pro	
	20	25		30	
Phe Gly Val Ser	Asn Cys Phe Leu	Leu Phe Ser	Val Cys Leu	Phe	
	35	40		45	
Pro Phe Cys Leu	Gly Ala Gly Ala	Gly Gly Glu	His Thr Ser	Tyr	
	50	55		60	
Leu His His Ser	Gly Leu Met Ser	Glu Gly Pro	Val Ser Pro	Ala	
	65	70		75	
Thr Tyr Leu Ala	Leu Ala Ser Thr	Ser Glu Arg	Leu Ile Thr	Ser	
	80	85		90	
Ser Pro His Ala	Gln Gly Cys Pro	Ser Gln Gly	Trp Leu Gly	Arg	
	95	100		105	
Ser His Gly Leu	Gly Pro Arg Arg	Ser Ser Gly	Leu Pro Pro	Gly	
	110	115		120	
Lys Ser Arg Ala	Ser Thr Ala Cys	Leu Gly Arg	Ala Pro Thr	Thr	
	125	130		135	
Arg His Gly Trp	Trp Leu Arg Leu	Lys Lys Ser	Leu Ser Met	Trp	
	140	145		150	
Glu Trp Glu Val	Leu Pro His Pro	Ala Trp Lys	Pro Arg Pro	Gly	
	155	160		165	
Ser Tyr Arg Gly	Leu Cys Asn Ser	Arg Gly Gly	His Met Lys	Met	
	170	175		180	
Glu Glu Pro Gly	Gly Ser Gly Ala	Pro Asp Val	Thr Ala Ser	Lys	
	185	190		195	
Ala Thr Gly Leu	Gly Arg Ala Ala	Pro Gln Glu	Gly		
	200	205			

&lt;210&gt; 8

&lt;211&gt; 291

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7494288CD1

&lt;400&gt; 8

Met	Leu	Arg	Ser	Pro	Thr	Phe	Thr	Asp	Ala	Gly	Pro	Arg	Cys	Ser	1	5	10	15
Cys	Leu	Pro	Val	Ser	Gln	Thr	Leu	Asp	Ser	Met	Asp	Thr	Val	Leu	20	25	30	35
Met	Gly	Ser	Leu	Gln	His	Cys	Cys	Cys	Leu	Leu	Pro	Lys	Met	Gly	40	45	50	55
Asp	Thr	Trp	Ala	Gln	Leu	Pro	Trp	Pro	Gly	Pro	Pro	His	Pro	Ala	60	65	70	75
Met	Leu	Leu	Ile	Ser	Leu	Leu	Leu	Ala	Ala	Gly	Leu	Met	His	Ser	80	85	90	95
Asp	Ala	Gly	Thr	Ser	Cys	Pro	Val	Leu	Cys	Thr	Cys	Arg	Asn	Gln	100	105	110	115
Val	Val	Asp	Cys	Ser	Ser	Gln	Arg	Leu	Phe	Ser	Val	Pro	Pro	Asp	120	125	130	135
Leu	Pro	Met	Asp	Thr	Arg	Asn	Leu	Ser	Leu	Ala	His	Asn	Arg	Ile	140	145	150	155
Thr	Ala	Val	Pro	Gly	Tyr	Leu	Thr	Cys	Tyr	Met	Glu	Leu	Gln		160	165	170	175
Val	Leu	Asp	Leu	His	Asn	Asn	Ser	Leu	Met	Glu	Leu	Pro	Arg	Gly	180	185	190	195
Leu	Phe	Leu	His	Ala	Lys	Arg	Leu	Ala	His	Leu	Asp	Leu	Ser	Tyr	200	205	210	215
Asn	Asn	Phe	Ser	His	Val	Pro	Ala	Asp	Met	Phe	Gln	Glu	Ala	His	220	225	230	235
Gly	Leu	Val	His	Ile	Asp	Leu	Ser	His	Asn	Pro	Trp	Leu	Arg	Arg	240	245	250	255
Val	His	Pro	Gln	Ala	Phe	Gln	Gly	Leu	Met	Gln	Leu	Arg	Asp	Leu	260	265	270	275
Asp	Leu	Ser	Tyr	Gly	Gly	Leu	Ala	Phe	Leu	Ser	Leu	Glu	Ala	Leu	280			
Glu	Gly	Leu	Pro	Gly	Leu	Val	Thr	Leu	Gln	Ile	Gly	Gly	Asn	Pro				
Trp	Val	Cys	Gly	Cys	Thr	Met	Glu	Pro	Leu	Leu	Lys	Trp	Leu	Arg				
Asn	Arg	Ile	Gln	Arg	Cys	Thr	Ala	Gly	Asn	Arg	Gly	Ala	Glu	Arg				
Gly	Ser	Gln	Gln	Gly	Gly	Leu	Ala	Ser	Met	Gly	Ser	Lys	Val	Ser				
Lys	Glu	Ser	Gly	Gly	Thr													

&lt;210&gt; 9

&lt;211&gt; 356

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474330CD1

&lt;400&gt; 9

Met	Pro	Ala	Ser	Ser	Leu	Pro	Gly	Lys	Leu	Trp	Phe	Val	Leu	Thr	1	5	10	15
Met	Leu	Leu	Arg	Met	Leu	Val	Ile	Val	Leu	Ala	Gly	Arg	Pro	Val	20	25	30	35
Tyr	Gln	Asp	Glu	Gln	Glu	Arg	Phe	Val	Cys	Asn	Thr	Leu	Gln	Pro	40	45	50	55
Gly	Cys	Ala	Asn	Val	Cys	Tyr	Asp	Val	Phe	Ser	Pro	Val	Ser	His				

Leu Arg Phe Trp Leu Ile Gln Gly Val Cys Val Leu Leu Pro Ser  
 65 70 75  
 Ala Val Phe Ser Val Tyr Val Leu His Arg Gly Ala Thr Leu Ala  
 80 85 90  
 Ala Leu Gly Pro Arg Arg Cys Pro Asp Pro Arg Glu Pro Ala Ser  
 95 100 105  
 Gly Gln Arg Arg Cys Pro Arg Pro Phe Gly Glu Arg Gly Gly Leu  
 110 115 120  
 Gln Val Pro Asp Phe Ser Ala Gly Tyr Ile Ile His Leu Leu Leu  
 125 130 135  
 Arg Thr Leu Leu Glu Ala Ala Phe Gly Ala Leu His Tyr Phe Leu  
 140 145 150  
 Phe Gly Phe Leu Ala Pro Lys Lys Phe Pro Cys Thr Arg Pro Pro  
 155 160 165  
 Cys Thr Gly Val Val Asp Cys Tyr Val Ser Arg Pro Thr Glu Lys  
 170 175 180  
 Ser Leu Leu Met Leu Phe Leu Trp Ala Val Ser Ala Leu Ser Phe  
 185 190 195  
 Leu Leu Gly Leu Ala Asp Leu Val Cys Ser Leu Arg Arg Arg Met  
 200 205 210  
 Arg Arg Arg Pro Gly Pro Pro Thr Ser Pro Ser Ile Arg Lys Gln  
 215 220 225  
 Ser Gly Ala Ser Gly His Ala Glu Gly Arg Arg Thr Asp Glu Glu  
 230 235 240  
 Gly Gly Arg Glu Glu Glu Gly Ala Pro Ala Pro Pro Gly Ala Arg  
 245 250 255  
 Ala Gly Gly Glu Gly Ala Gly Ser Pro Arg Arg Thr Ser Arg Val  
 260 265 270  
 Ser Gly His Thr Lys Ile Pro Asp Glu Asp Glu Ser Glu Val Thr  
 275 280 285  
 Ser Ser Ala Ser Glu Lys Leu Gly Arg Gln Pro Arg Gly Arg Pro  
 290 295 300  
 His Arg Glu Ala Ala Gln Asp Pro Arg Gly Ser Gly Ser Glu Glu  
 305 310 315  
 Gln Pro Ser Ala Ala Pro Ser Arg Leu Ala Ala Pro Pro Ser Cys  
 320 325 330  
 Ser Ser Leu Gln Pro Pro Asp Pro Pro Ala Ser Ser Ser Gly Ala  
 335 340 345  
 Pro His Leu Arg Ala Arg Lys Ser Glu Trp Val  
 350 355

<210> 10  
 <211> 82  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5911370CD1

<400> 10  
 Met Glu Leu Ile Lys Ser Arg Ala Thr Val Cys Ala Leu Leu Leu  
 1 5 10 15  
 Ala Leu Leu Leu Leu Ser His Tyr Asp Gly Gly Thr Thr Thr Thr  
 20 25 30  
 Met Val Ala Glu Ala Arg Val Cys Met Gly Lys Ser Gln His His  
 35 40 45  
 Ser Phe Pro Cys Ile Ser Asp Arg Leu Cys Ser Asn Glu Cys Val  
 50 55 60  
 Lys Glu Asp Gly Gly Trp Thr Ala Gly Tyr Cys His Leu Arg Tyr  
 65 70 75  
 Cys Arg Cys Gln Lys Ala Cys  
 80



<210> 11  
 <211> 529  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7647134CD1

<400> 11  
 Met Arg Pro Gln Cys Thr Pro Ala His Arg Pro Gly Ala Ile Leu  
 1 5 10 15  
 Leu Thr Ala Gln Pro Arg Ala Pro Thr Val Ala Ser Cys Gly Tyr  
 20 25 30  
 Cys Ser Cys Lys Leu Arg Pro Thr Arg Arg Ser Pro Thr Gly Lys  
 35 40 45  
 Ala Val Val Arg Pro Pro Pro Gly Ala Pro Gln Gln Pro Gly  
 50 55 60  
 Val His Ser Ser Ser Leu Pro His Arg Pro Ser Leu Phe Ser Ser  
 65 70 75  
 Ser Pro Glu Val Glu Phe Glu Arg Arg His Gly Gly Gly Ala Ala  
 80 85 90  
 Leu Leu Pro Leu Leu Cys Leu Pro Met Asp Met His Cys Lys Ala  
 95 100 105  
 Asp Pro Phe Ser Ala Met His Pro Gly His Gly Gly Val Asn Gln  
 110 115 120  
 Leu Gly Gly Val Phe Val Asn Gly Arg Pro Leu Pro Asp Val Val  
 125 130 135  
 Arg Gln Arg Ile Val Glu Leu Ala His Gln Gly Val Arg Pro Cys  
 140 145 150  
 Asp Ile Ser Arg Gln Leu Arg Val Ser His Gly Cys Val Ser Lys  
 155 160 165  
 Ile Leu Gly Arg Tyr Tyr Glu Thr Gly Ser Ile Lys Pro Gly Val  
 170 175 180  
 Ile Gly Gly Ser Lys Pro Lys Val Ala Thr Pro Lys Val Val Asp  
 185 190 195  
 Lys Ile Ala Glu Tyr Lys Arg Gln Asn Pro Thr Met Phe Ala Trp  
 200 205 210  
 Glu Ile Arg Asp Arg Leu Leu Ala Glu Gly Ile Cys Asp Asn Asp  
 215 220 225  
 Thr Val Pro Ser Val Ser Ser Ile Asn Arg Ile Ile Arg Thr Lys  
 230 235 240  
 Val Gln Gln Pro Phe His Pro Thr Pro Asp Gly Ala Gly Thr Gly  
 245 250 255  
 Val Thr Ala Pro Gly His Thr Ile Val Pro Ser Thr Ala Ser Pro  
 260 265 270  
 Pro Val Ser Ser Ala Ser Asn Asp Pro Val Gly Ser Tyr Ser Ile  
 275 280 285  
 Asn Gly Ile Leu Gly Ile Pro Arg Ser Asn Gly Glu Lys Arg Lys  
 290 295 300  
 Arg Asp Glu Val Glu Val Tyr Thr Asp Pro Ala His Ile Arg Gly  
 305 310 315  
 Gly Gly Gly Leu His Leu Val Trp Thr Leu Arg Asp Val Ser Glu  
 320 325 330  
 Gly Ser Val Pro Asn Gly Asp Ser Gln Ser Gly Val Asp Ser Leu  
 335 340 345  
 Arg Lys His Leu Arg Ala Asp Thr Phe Thr Gln Gln Gln Leu Glu  
 350 355 360  
 Ala Leu Asp Arg Val Phe Glu Arg Pro Ser Tyr Pro Asp Val Phe  
 365 370 375  
 Gln Ala Ser Glu His Ile Lys Ser Glu Gln Gly Asn Glu Tyr Ser  
 380 385 390  
 Leu Pro Ala Leu Thr Pro Gly Leu Asp Glu Val Lys Ser Ser Leu

	395		400		405
Ser Ala Ser Thr	Asn Pro Glu Leu Gly	Ser Asn Val Ser Gly Thr			
	410		415		420
Gln Thr Tyr Pro	Val Val Thr Gly Arg	Asp Met Ala Ser Thr Thr			
	425		430		435
Leu Pro Gly Tyr	Pro Pro His Val Pro	Pro Thr Gly Gln Gly Ser			
	440		445		450
Tyr Pro Thr Ser	Thr Leu Ala Gly Met	Val Pro Gly Ser Glu Phe			
	455		460		465
Ser Gly Asn Pro	Tyr Ser His Pro Gln	Tyr Thr Ala Tyr Asn Glu			
	470		475		480
Ala Trp Arg Phe	Ser Asn Pro Ala Leu	Leu Met Pro Pro Pro Gly			
	485		490		495
Pro Pro Leu Pro	Leu Val Pro Leu Pro	Met Thr Ala Thr Ser Tyr			
	500		505		510
Arg Gly Asp His	Ile Lys Leu Gln Ala	Asp Ser Phe Gly Leu His			
	515		520		525
Ile Val Pro Val					

<210> 12  
 <211> 453  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1631327CD1

<400> 12

Met. Gly Val Leu Gly Arg Val Leu Leu Trp	Leu Gln Leu Cys Ala
1	5 10 15
Leu Thr Gln Ala Val Ser Lys Leu Trp Val	Pro Asn Thr Asp Phe
	20 25 30
Asp Val Ala Ala Asn Trp Ser Gln Asn Arg	Thr Pro Cys Ala Gly
	35 40 45
Gly Ala Val Glu Phe Pro Ala Asp Lys Met	Val Ser Val Leu Val
	50 55 60
Gln Glu Gly His Ala Val Ser Asp Met Leu	Leu Pro Leu Asp Gly
	65 70 75
Glu Leu Val Leu Ala Ser Gly Ala Gly Phe	Gly Val Ser Asp Val
	80 85 90
Gly Ser His Leu Asp Cys Gly Ala Gly Glu	Pro Ala Val Phe Arg
	95 100 105
Asp Ser Asp Arg Phe Ser Trp His Asp Pro	His Leu Trp Arg Ser
	110 115 120
Gly Asp Glu Ala Pro Gly Leu Phe Phe Val	Asp Ala Glu Arg Val
	125 130 135
Pro Cys Arg His Asp Asp Val Phe Phe Pro	Pro Ser Ala Ser Phe
	140 145 150
Arg Val Gly Leu Gly Pro Gly Ala Ser Pro	Val Arg Val Arg Ser
	155 160 165
Ile Ser Ala Leu Gly Arg Thr Phe Thr Arg	Asp Glu Asp Leu Ala
	170 175 180
Val Phe Leu Ala Ser Arg Ala Gly Arg Leu	Arg Phe His Gly Pro
	185 190 195
Gly Ala Leu Ser Val Gly Pro Glu Asp Cys	Ala Asp Pro Ser Gly
	200 205 210
Cys Val Cys Gly Asn Ala Glu Ala Gln Pro	Trp Ile Cys Ala Ala
	215 220 225
Leu Leu Gln Pro Leu Gly Gly Arg Cys Pro	Gln Ala Ala Cys His
	230 235 240
Ser Ala Leu Arg Pro Gln Gly Gln Cys Cys	Asp Leu Cys Gly Ala

Val Val Leu Leu	Thr His Gly Pro Ala	Phe Asp Leu Glu Arg Tyr	245	250	255
Arg Ala Arg Ile	Leu Asp Thr Phe Leu	Gly Leu Pro Gln Tyr His	260	265	270
Gly Leu Gln Val	Ala Val Ser Lys Val	Pro Arg Ser Ser Arg Leu	275	280	285
Arg Glu Ala Asp	Thr Glu Ile Gln Val	Val Leu Val Glu Asn Gly	290	295	300
Pro Glu Thr Gly	Gly Ala Gly Arg Leu	Ala Arg Ala Leu Leu Ala	305	310	315
Asp Val Ala Glu	Asn Gly Glu Ala Leu	Gly Val Leu Glu Ala Thr	320	325	330
Met Arg Glu Ser	Gly Ala His Val Trp	Gly Ser Ser Ala Ala Gly	335	340	345
Leu Ala Gly Gly	Val Ala Ala Ala Val	Leu Leu Ala Leu Leu Val	350	355	360
Leu Leu Val Ala	Pro Pro Leu Leu Arg	Arg Ala Gly Arg Leu Arg	365	370	375
Trp Arg Arg His	Glu Ala Ala Ala Pro	Ala Gly Ala Pro Leu Gly	380	385	390
Phe Arg Asn Pro	Val Phe Asp Val Thr	Ala Ser Glu Glu Leu Pro	395	400	405
Leu Pro Arg Arg	Leu Ser Leu Val Pro	Lys Ala Ala Ala Asp Ser	410	415	420
Thr Ser His Ser	Tyr Phe Val Asn Pro	Leu Phe Ala Gly Ala Glu	425	430	435
Ala Glu Ala			440	445	450

<210> 13  
 <211> 271  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 044232CD1

<400> 13

Met Ala Ala Gly Gly	Arg Met Glu Asp Gly	Ser Leu Asp Ile Thr	1	5	10	15
Gln Ser Ile Glu Asp	Asp Pro Leu Leu Asp	Ala Gln Leu Leu Pro	20	25	30	35
His His Ser Leu Gln	Ala His Phe Arg Pro	Arg Phe His Pro Leu	40	45	50	55
Pro Thr Val Ile Ile	Val Asn Leu Leu Trp	Phe Ile His Leu Val	60	65	70	75
Phe Val Val Leu Ala	Phe Leu Thr Gly Val	Leu Cys Ser Tyr Pro	80	85	90	95
Asn Pro Asn Glu Asp	Lys Cys Pro Gly Asn	Tyr Thr Asn Pro Leu	100	105	110	115
Lys Val Gln Thr Val	Ile Ile Leu Gly Lys	Val Ile Leu Trp Ile	120	125	130	135
Leu His Leu Leu Leu	Glu Cys Tyr Ile Gln	Tyr His His Ser Lys	140	145	150	155
Ile Arg Asn Arg Gly	Tyr Asn Leu Ile Tyr	Arg Ser Thr Arg His	160	165	170	175
Leu Lys Arg Leu Ala	Leu Met Ile Gln Ser	Ser Gly Asn Thr Val	180	185	190	195
Leu Leu Leu Ile Leu	Cys Met Gln His Ser	Phe Pro Glu Pro Gly	200	205	210	215
Arg Leu Tyr Leu Asp	Leu Ile Leu Ala Ile	Leu Ala Leu Glu Leu	220	225	230	235

	170		175		180
Ile Cys Ser Leu	Ile Cys Leu Leu Ile Tyr Thr Val Lys Ile Arg				
	185		190		195
Arg Phe Asn Lys	Ala Lys Pro Glu Pro Asp Ile Leu Glu Glu Glu				
	200		205		210
Lys Ile Tyr Ala	Tyr Pro Ser Asn Ile Thr Ser Glu Thr Gly Phe				
	215		220		225
Arg Thr Ile Ser	Ser Leu Glu Glu Ile Val Glu Lys Gln Gly Asp				
	230		235		240
Thr Ile Glu Tyr	Leu Lys Arg His Asn Ala Leu Leu Ser Lys Arg				
	245		250		255
Leu Leu Ala Leu	Thr Ser Ser Asp Leu Gly Cys Gln Pro Ser Arg				
	260		265		270
Thr					

<210> 14  
 <211> 203  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 560293CD1

<400> 14	
Met Ala Cys Gly Ala Thr Leu Lys Arg Thr Leu Asp Phe Asp Pro	
1 5 10 15	
Leu Leu Ser Pro Ala Ser Pro Lys Arg Arg Arg Cys Ala Pro Leu	
20 25 30	
Ser Ala Pro Thr Ser Ala Ala Ala Ser Pro Leu Ser Ala Ala Ala	
35 40 45	
Ala Thr Ala Ala Ser Phe Ser Ala Ala Ala Ala Ser Pro Gln Lys	
50 55 60	
Tyr Leu Arg Met Glu Pro Ser Pro Phe Gly Asp Val Ser Ser Arg	
65 70 75	
Leu Thr Thr Glu Gln Ile Leu Tyr Asn Ile Lys Gln Glu Tyr Lys	
80 85 90	
Arg Met Gln Lys Arg Arg His Leu Glu Thr Ser Phe Gln Gln Thr	
95 100 105	
Asp Pro Cys Cys Thr Ser Asp Ala Gln Pro His Ala Phe Leu Leu	
110 115 120	
Ser Gly Pro Ala Ser Pro Gly Thr Ser Ser Ala Ala Ser Ser Pro	
125 130 135	
Leu Lys Lys Glu Gln Pro Leu Phe Thr Leu Arg Gln Val Gly Met	
140 145 150	
Ile Cys Glu Arg Leu Leu Lys Glu Arg Glu Glu Lys Val Arg Glu	
155 160 165	
Glu Tyr Glu Glu Ile Leu Asn Thr Lys Leu Ala Glu Gln Tyr Asp	
170 175 180	
Ala Phe Val Lys Phe Thr His Asp Gln Ile Met Arg Arg Tyr Gly	
185 190 195	
Glu Gln Pro Ala Ser Tyr Val Ser	
200	

<210> 15  
 <211> 529  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2025618CD1

<400> 15  
Met Ala Ala Leu Thr Thr Val Val Val Ala Ala Ala Ala Thr Ala  
1 5 10 15  
Val Ala Gly Ala Val Ala Gly Ala Gly Ala Ala Thr Gly Thr Gly  
20 25 30  
Val Gly Ala Thr Pro Ala Pro Gln Gln Ser Asp Gly Cys Phe Ser  
35 40 45  
Thr Ser Gly Gly Ile Arg Pro Phe His Leu Gln Asn Trp Lys Gln  
50 55 60  
Lys Val Asn Gln Thr Lys Lys Ala Glu Phe Val Arg Thr Ala Glu  
65 70 75  
Lys Phe Lys Asn Gln Val Ile Asn Met Glu Lys Asp Lys His Ser  
80 85 90  
His Phe Tyr Asn Gln Lys Ser Asp Phe Arg Ile Glu His Ser Met  
95 100 105  
Leu Glu Glu Leu Glu Asn Lys Leu Ile His Ser Arg Lys Thr Glu  
110 115 120  
Arg Ala Lys Ile Gln Gln Gln Leu Ala Lys Ile His Asn Asn Val  
125 130 135  
Lys Lys Leu Gln His Gln Leu Lys Asp Val Lys Pro Thr Pro Asp  
140 145 150  
Phe Val Glu Lys Leu Arg Glu Met Met Glu Glu Ile Glu Asn Ala  
155 160 165  
Ile Asn Thr Phe Lys Glu Glu Gln Arg Leu Ile Tyr Glu Glu Leu  
170 175 180  
Ile Lys Glu Glu Lys Thr Thr Asn Asn Glu Leu Ser Ala Ile Ser  
185 190 195  
Arg Lys Ile Asp Thr Trp Ala Leu Gly Asn Ser Glu Thr Glu Lys  
200 205 210  
Ala Phe Arg Ala Ile Ser Ser Lys Val Pro Val Asp Lys Val Thr  
215 220 225  
Pro Ser Thr Leu Pro Glu Glu Val Leu Asp Phe Glu Lys Phe Leu  
230 235 240  
Gln Gln Thr Gly Gly Arg Gln Gly Ala Trp Asp Asp Tyr Asp His  
245 250 255  
Gln Asn Phe Val Lys Val Arg Asn Lys His Lys Gly Lys Pro Thr  
260 265 270  
Phe Met Glu Glu Val Leu Glu His Leu Pro Gly Lys Thr Gln Asp  
275 280 285  
Glu Val Gln Gln His Glu Lys Trp Tyr Gln Lys Phe Leu Ala Leu  
290 295 300  
Glu Glu Arg Lys Lys Glu Ser Ile Gln Ile Trp Lys Thr Lys Lys  
305 310 315  
Gln Gln Lys Arg Glu Glu Ile Phe Lys Leu Lys Glu Lys Ala Asp  
320 325 330  
Asn Thr Pro Val Leu Phe His Asn Lys Gln Glu Asp Asn Gln Lys  
335 340 345  
Gln Lys Glu Glu Gln Arg Lys Lys Gln Lys Leu Ala Val Glu Ala  
350 355 360  
Trp Lys Lys Gln Lys Ser Ile Glu Met Ser Met Lys Cys Ala Ser  
365 370 375  
Gln Leu Lys Glu Glu Glu Lys Glu Lys Lys His Gln Lys Glu  
380 385 390  
Arg Gln Arg Gln Phe Lys Leu Lys Leu Leu Leu Glu Ser Tyr Thr  
395 400 405  
Gln Gln Lys Lys Glu Gln Glu Glu Phe Leu Arg Leu Glu Lys Glu  
410 415 420  
Ile Arg Glu Lys Ala Glu Lys Ala Glu Lys Arg Lys Asn Ala Ala  
425 430 435  
Asp Glu Ile Ser Arg Phe Gln Glu Arg Asp Leu His Lys Leu Glu  
440 445 450  
Leu Lys Ile Leu Asp Arg Gln Ala Lys Glu Asp Glu Lys Ser Gln  
455 460 465

Lys	Gln	Arg	Arg	Leu	Ala	Lys	Leu	Lys	Glu	Lys	Val	Glu	Asn	Asn
				470					475					480
Val	Ser	Arg	Asp	Pro	Ser	Arg	Leu	Tyr	Lys	Pro	Thr	Lys	Gly	Trp
				485					490					495
Glu	Glu	Arg	Thr	Lys	Lys	Ile	Gly	Pro	Thr	Gly	Ser	Gly	Pro	Leu
				500					505					510
Leu	His	Ile	Pro	His	Arg	Ala	Ile	Pro	Thr	Trp	Arg	Gln	Gly	Ile
				515					520					525
Gln	Arg	Arg	Val											

&lt;210&gt; 16

&lt;211&gt; 305

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3342443CD1

&lt;400&gt; 16

Met	Lys	Ala	Leu	Gly	Ala	Val	Leu	Leu	Ala	Leu	Leu	Leu	Cys	Gly
1				5					10					15
Arg	Pro	Gly	Arg	Gly	Gln	Thr	Gln	Gln	Glu	Glu	Glu	Glu	Glu	Asp
				20					25					30
Glu	Asp	His	Gly	Pro	Asp	Asp	Tyr	Asp	Glu	Glu	Asp	Glu	Asp	Glu
				35					40					45
Val	Glu	Glu	Glu	Glu	Thr	Asn	Arg	Leu	Pro	Gly	Gly	Arg	Ser	Arg
				50					55					60
Val	Leu	Leu	Arg	Cys	Tyr	Thr	Cys	Lys	Ser	Leu	Pro	Arg	Asp	Glu
				65					70					75
Arg	Cys	Asn	Leu	Thr	Gln	Asn	Cys	Ser	His	Gly	Gln	Thr	Cys	Thr
				80					85					90
Thr	Leu	Ile	Ala	His	Gly	Asn	Thr	Glu	Ser	Gly	Leu	Leu	Thr	Thr
				95					100					105
His	Ser	Thr	Trp	Cys	Thr	Asp	Ser	Cys	Gln	Pro	Ile	Thr	Lys	Thr
				110					115					120
Val	Glu	Gly	Thr	Gln	Val	Thr	Met	Thr	Cys	Cys	Gln	Ser	Ser	Leu
				125					130					135
Cys	Asn	Val	Pro	Pro	Trp	Gln	Ser	Ser	Arg	Val	Gln	Asp	Pro	Thr
				140					145					150
Gly	Lys	Gly	Ala	Gly	Gly	Pro	Arg	Gly	Ser	Ser	Glu	Thr	Val	Gly
				155					160					165
Ala	Ala	Pro	Ala	Gln	Pro	Pro	Cys	Arg	Pro	Trp	Ser	Asn	Gly	Gly
				170					175					180
Gln	Glu	Thr	Leu	Thr	His	Gly	Pro	Ser	Pro	Pro	Pro	Pro	Gly	Ser
				185					190					195
Pro	Pro	Ala	Leu	Pro	Ala	Leu	Cys	Leu	Val	Pro	Ser	Pro	Pro	Ala
				200					205					210
Pro	Ala	Pro	Ala	Leu	Glu	Asn	Gly	Phe	Gly	Val	Ser	Trp	Ala	Ile
				215					220					225
Gln	Pro	Ala	Gln	Ala	Pro	Arg	Pro	Gly	Cys	Phe	Leu	Ser	Ser	Arg
				230					235					240
Leu	Cys	Pro	Trp	Cys	Pro	Phe	Ser	Thr	Thr	Cys	Glu	Gln	Gln	Asp
				245					250					255
Cys	Arg	Thr	Trp	Ala	Pro	Gly	Ser	Arg	Pro	Arg	Leu	Ala	Arg	Pro
				260					265					270
Arg	Ala	Leu	Gln	Pro	Ser	Arg	Gly	Ala	Gly	Gly	Ser	His	Gln	His
				275					280					285
Ser	Gln	Ala	Glu	Met	Ile	Pro	Pro	His	Ser	Trp	Gly	Pro	Pro	His
				290					295					300
Pro	Val	Leu	Thr	Pro										
				305										

<210> 17  
 <211> 493  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2267957CD1

<400> 17  
 Met His Pro His Arg Asp Pro Arg Gly Leu Trp Leu Leu Leu Pro  
 1 5 10 15  
 Ser Leu Ser Leu Leu Leu Phe Glu Val Ala Arg Ala Gly Arg Ala  
 20 25 30  
 Val Val Ser Cys Pro Ala Ala Cys Leu Cys Ala Ser Asn Ile Leu  
 35 40 45  
 Ser Cys Ser Lys Gln Gln Leu Pro Asn Val Pro His Ser Leu Pro  
 50 55 60  
 Ser Tyr Thr Ala Leu Leu Asp Leu Ser His Asn Asn Leu Ser Arg  
 65 70 75  
 Leu Arg Ala Glu Trp Thr Pro Thr Arg Leu Thr Gln Leu His Ser  
 80 85 90  
 Leu Leu Leu Ser His Asn His Leu Asn Phe Ile Ser Ser Glu Ala  
 95 100 105  
 Phe Ser Pro Val Pro Asn Leu Arg Tyr Leu Asp Leu Ser Ser Asn  
 110 115 120  
 Gln Leu Arg Thr Leu Asp Glu Phe Leu Phe Ser Asp Leu Gln Val  
 125 130 135  
 Leu Glu Val Leu Leu Leu Tyr Asn Asn His Ile Met Ala Val Asp  
 140 145 150  
 Arg Cys Ala Phe Asp Asp Met Ala Gln Leu Gln Lys Leu Tyr Leu  
 155 160 165  
 Ser Gln Asn Gln Ile Ser Arg Phe Pro Leu Glu Leu Val Lys Glu  
 170 175 180  
 Gly Ala Lys Leu Pro Lys Leu Thr Leu Leu Asp Leu Ser Ser Asn  
 185 190 195  
 Lys Leu Lys Asn Leu Pro Leu Pro Asp Leu Gln Lys Leu Pro Ala  
 200 205 210  
 Trp Ile Lys Asn Gly Leu Tyr Leu His Asn Asn Pro Leu Asn Cys  
 215 220 225  
 Asp Cys Glu Leu Tyr Gln Leu Phe Ser His Trp Gln Tyr Arg Gln  
 230 235 240  
 Leu Ser Ser Val Met Asp Phe Gln Glu Asp Leu Tyr Cys Met Asn  
 245 250 255  
 Ser Lys Lys Leu His Asn Val Phe Asn Leu Ser Phe Leu Asn Cys  
 260 265 270  
 Gly Glu Tyr Lys Glu Arg Ala Trp Glu Ala His Leu Gly Asp Thr  
 275 280 285  
 Leu Ile Ile Lys Cys Asp Thr Lys Gln Gln Gly Met Thr Lys Val  
 290 295 300  
 Trp Val Thr Pro Ser Asn Glu Arg Val Leu Asp Glu Val Thr Asn  
 305 310 315  
 Gly Thr Val Ser Val Ser Lys Asp Gly Ser Leu Leu Phe Gln Gln  
 320 325 330  
 Val Gln Val Glu Asp Gly Gly Val Tyr Thr Cys Tyr Ala Met Gly  
 335 340 345  
 Glu Thr Phe Asn Glu Thr Leu Ser Val Glu Leu Lys Val His Asn  
 350 355 360  
 Phe Thr Leu His Gly His His Asp Thr Leu Asn Thr Ala Tyr Thr  
 365 370 375  
 Thr Leu Val Gly Cys Ile Leu Ser Val Val Leu Val Leu Ile Tyr  
 380 385 390  
 Leu Tyr Leu Thr Pro Cys Arg Cys Trp Cys Arg Gly Val Glu Lys

	395		400		405
Pro Ser Ser His	Gln Gly Asp Ser Leu	Ser Ser Ser Met Leu	Ser		
	410		415		420
Thr Thr Pro Asn	His Asp Pro Met Ala	Gly Gly Asp Lys Asp	Asp		
	425		430		435
Gly Phe Asp Arg	Arg Val Ala Phe Leu	Glu Pro Ala Gly Pro	Gly		
	440		445		450
Gln Gly Gln Asn	Gly Lys Leu Lys Pro	Gly Asn Thr Leu Pro	Val		
	455		460		465
Pro Glu Ala Thr	Gly Lys Gly Gln Arg	Arg Met Ser Asp Pro	Glu		
	470		475		480
Ser Val Ser Ser	Val Phe Ser Asp Thr	Pro Ile Val Val			
	485		490		

&lt;210&gt; 18

&lt;211&gt; 869

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480277CD1

&lt;400&gt; 18

Met Glu Val Thr	Cys Leu Leu Leu Leu	Ala Leu Ile Pro Phe	His
1	5	10	15
Cys Arg Gly Gln	Gly Val Tyr Ala Pro	Ala Gln Ala Gln Ile	Val
	20	25	30
His Ala Gly Gln	Ala Cys Val Val Lys	Glu Asp Asn Ile Ser	Glu
	35	40	45
Arg Val Tyr Thr	Ile Arg Glu Gly Asp	Thr Leu Met Leu Gln	Cys
	50	55	60
Leu Val Thr Gly	His Pro Arg Pro Gln	Val Arg Trp Thr Lys	Thr
	65	70	75
Ala Gly Ser Ala	Ser Asp Lys Phe Gln	Glu Thr Ser Val Phe	Asn
	80	85	90
Glu Thr Leu Arg	Ile Glu Arg Ile Ala	Arg Thr Gln Gly Gly	Arg
	95	100	105
Tyr Tyr Cys Lys	Ala Glu Asn Gly Val	Gly Val Pro Ala Ile	Lys
	110	115	120
Ser Ile Arg Val	Asp Val Gln Ser Met	Lys Asn Ala Thr Phe	Gln
	125	130	135
Ile Thr Pro Asp	Val Ile Lys Glu Ser	Glu Asn Ile Gln Leu	Gly
	140	145	150
Gln Asp Leu Lys	Leu Ser Cys His Val	Asp Ala Val Pro Gln	Glu
	155	160	165
Lys Val Thr Tyr	Gln Trp Phe Lys Asn	Gly Lys Pro Ala Arg	Met
	170	175	180
Ser Lys Arg Leu	Leu Val Thr Arg Asn	Asp Pro Glu Leu Pro	Ala
	185	190	195
Val Thr Ser Ser	Leu Glu Leu Ile Asp	Leu His Phe Ser Asp	Tyr
	200	205	210
Gly Thr Tyr Leu	Cys Met Ala Ser Phe	Pro Gly Ala Pro Val	Pro
	215	220	225
Asp Leu Ser Val	Glu Val Asn Ile Ser	Ser Glu Thr Val Pro	Pro
	230	235	240
Thr Ile Ser Val	Pro Lys Gly Arg Ala	Val Val Thr Val Arg	Glu
	245	250	255
Gly Ser Pro Ala	Glu Leu Gln Cys Glu	Val Arg Gly Lys Pro	Arg
	260	265	270
Pro Pro Val Leu	Trp Ser Arg Val Asp	Lys Glu Ala Ala Leu	Leu
	275	280	285
Pro Ser Gly Leu	Pro Leu Glu Glu Thr	Pro Asp Gly Lys Leu	Arg



	290		295		300
Leu Glu Arg Val	Ser Arg Asp Met Ser	Gly Thr Tyr Arg Cys Gln			
	305		310		315
Thr Ala Arg Tyr	Asn Gly Phe Asn Val	Arg Pro Arg Glu Ala Gln			
	320		325		330
Val Gln Leu Asn	Val Gln Phe Pro Pro	Glu Val Glu Pro Ser Ser			
	335		340		345
Gln Asp Val Arg	Gln Ala Leu Gly Arg	Pro Val Leu Leu Arg Cys			
	350		355		360
Ser Leu Leu Arg	Gly Ser Pro Gln Arg	Ile Ala Ser Ala Val Trp			
	365		370		375
Arg Phe Lys Gly	Gln Leu Leu Pro Pro	Pro Pro Val Val Pro Ala			
	380		385		390
Ala Ala Glu Ala	Pro Asp His Ala Glu	Leu Arg Leu Asp Ala Val			
	395		400		405
Thr Arg Asp Ser	Ser Gly Ser Tyr Glu	Cys Ser Val Ser Asn Asp			
	410		415		420
Val Gly Ser Ala	Ala Cys Leu Phe Gln	Val Ser Ala Lys Ala Tyr			
	425		430		435
Ser Pro Glu Phe	Tyr Phe Asp Thr Pro	Asn Pro Thr Arg Ser His			
	440		445		450
Lys Leu Ser Lys	Asn Tyr Ser Tyr Val	Leu Gln Trp Thr Gln Arg			
	455		460		465
Glu Pro Asp Ala	Val Asp Pro Val Leu	Asn Tyr Arg Leu Ser Ile			
	470		475		480
Arg Gln Leu Asn	Gln His Asn Ala Val	Val Lys Ala Ile Pro Val			
	485		490		495
Arg Arg Val Glu	Lys Gly Gln Leu Leu	Glu Tyr Ile Leu Thr Asp			
	500		505		510
Leu Arg Val Pro	His Ser Tyr Glu Val	Arg Leu Thr Pro Tyr Thr			
	515		520		525
Thr Phe Gly Ala	Gly Asp Met Ala Ser	Arg Ile Ile His Tyr Thr			
	530		535		540
Glu Arg Gln Ile	Arg Trp Pro Pro Val	Leu Ala Leu Arg Thr Leu			
	545		550		555
Ser Ser Gly Pro	Lys Gln Gly Ile Leu	Cys Arg Ala Pro His Leu			
	560		565		570
Ser Ser Asp Leu	Val Ser Pro Leu Ala	Phe Ser Ala Ile Asn Ser			
	575		580		585
Pro Asn Leu Ser	Asp Asn Thr Cys His	Phe Glu Asp Glu Lys Ile			
	590		595		600
Cys Gly Tyr Thr	Gln Asp Leu Thr Asp	Asn Phe Asp Trp Thr Arg			
	605		610		615
Gln Asn Ala Leu	Thr Gln Asn Pro Lys	Arg Ser Pro Asn Thr Gly			
	620		625		630
Pro Pro Thr Asp	Ile Ser Gly Thr Pro	Glu Gly Tyr Tyr Met Phe			
	635		640		645
Ile Glu Thr Ser	Arg Pro Arg Glu Leu	Gly Asp Arg Ala Arg Leu			
	650		655		660
Val Ser Pro Leu	Tyr Asn Ala Ser Ala	Lys Phe Tyr Cys Val Ser			
	665		670		675
Phe Phe Tyr His	Met Tyr Gly Lys His	Ile Gly Ser Leu Asn Leu			
	680		685		690
Leu Val Arg Ser	Arg Asn Lys Gly Ala	Leu Asp Thr His Ala Trp			
	695		700		705
Ser Leu Ser Gly	Asn Lys Gly Asn Val	Trp Gln Gln Ala His Val			
	710		715		720
Pro Ile Ser Pro	Ser Gly Pro Phe Gln	Ile Ile Phe Glu Gly Val			
	725		730		735
Arg Gly Pro Gly	Tyr Leu Gly Asp Ile	Ala Ile Asp Asp Val Thr			
	740		745		750
Leu Lys Lys Gly	Glu Cys Pro Arg Lys	Gln Thr Asp Pro Asn Lys			
	755		760		765

```
<210> 19
<211> 174
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> misc_feature  
<223> Incyte ID No: 3450647CD1
```

```
<210> 20
<211> 561
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> misc_feature  
<223> Incyte ID No: 2053428CD1
```

<400> 20  
Met Cys Leu Phe Leu Pro Val Leu Ser Ser Glu Ser Ala Pro Leu  
1 5 10 15  
Val Val Arg Lys Gly Ser Asp Val Val Ala Gly Lys Met Ala Thr

Ala Ala Thr Ile	Pro Ser Val Ala Thr	Ala Thr Ala Ala Ala	Leu
35	40		45
Gly Glu Val Glu	Asp Glu Gly Leu Leu	Ala Ser Leu Phe Arg	Asp
50	55		60
Arg Phe Pro Glu	Ala Gln Trp Arg Glu	Arg Pro Asp Val Gly	Arg
65	70		75
Tyr Leu Arg Glu	Leu Ser Gly Ser Gly	Leu Glu Arg Leu Arg	Arg
80	85		90
Glu Pro Glu Arg	Leu Ala Glu Glu Arg	Ala Gln Leu Leu Gln	Gln
95	100		105
Thr Arg Asp Leu	Ala Phe Ala Asn Tyr	Lys Thr Phe Ile Arg	Gly
110	115		120
Ala Glu Cys Thr	Glu Arg Ile His Arg	Leu Phe Gly Asp Val	Glu
125	130		135
Ala Ser Leu Gly	Arg Leu Leu Asp Arg	Leu Pro Ser Phe Gln	Gln
140	145		150
Ser Cys Arg Asn	Phe Val Lys Glu Ala	Glu Glu Ile Ser Ser	Asn
155	160		165
Arg Arg Met Asn	Ser Leu Thr Leu Asn	Arg His Thr Glu Ile	Leu
170	175		180
Glu Ile Leu Glu	Ile Pro Gln Leu Met	Asp Thr Cys Val Arg	Asn
185	190		195
Ser Tyr Tyr Glu	Glu Ala Leu Glu Leu	Ala Ala Tyr Val Arg	Arg
200	205		210
Leu Glu Arg Lys	Tyr Ser Ser Ile Pro	Val Ile Gln Gly Ile	Val
215	220		225
Asn Glu Val Arg	Gln Ser Met Gln Leu	Met Leu Ser Gln Leu	Ile
230	235		240
Gln Gln Leu Arg	Thr Asn Ile Gln Leu	Pro Ala Cys Leu Arg	Val
245	250		255
Ile Gly Tyr Leu	Arg Arg Met Asp Val	Phe Thr Glu Ala Glu	Leu
260	265		270
Arg Val Lys Phe	Leu Gln Ala Arg Asp	Ala Trp Leu Arg Ser	Ile
275	280		285
Leu Thr Ala Ile	Pro Asn Asp Asp Pro	Tyr Phe His Ile Thr	Lys
290	295		300
Thr Ile Glu Ala	Ser Arg Val His Leu	Phe Asp Ile Ile Thr	Gln
305	310		315
Tyr Arg Ala Ile	Phe Ser Asp Glu Asp	Pro Leu Leu Pro Pro	Ala
320	325		330
Met Gly Glu His	Thr Val Asn Glu Ser	Ala Ile Phe His Gly	Trp
335	340		345
Val Leu Gln Lys	Val Ser Gln Phe Leu	Gln Val Leu Glu Thr	Asp
350	355		360
Leu Tyr Arg Gly	Ile Gly Gly His Leu	Asp Ser Leu Leu Gly	Gln
365	370		375
Cys Met Tyr Phe	Gly Leu Ser Phe Ser	Arg Val Gly Ala Asp	Phe
380	385		390
Arg Gly Gln Leu	Ala Pro Val Phe Gln	Arg Val Ala Ile Ser	Thr
395	400		405
Phe Gln Lys Ala	Ile Gln Glu Thr Val	Glu Lys Phe Gln Glu	Glu
410	415		420
Met Asn Ser Tyr	Met Leu Ile Ser Ala	Pro Ala Ile Leu Gly	Thr
425	430		435
Ser Asn Met Pro	Ala Ala Val Pro Ala	Thr Gln Pro Gly Thr	Leu
440	445		450
Gln Pro Pro Met	Val Leu Leu Asp Phe	Pro Pro Leu Ala Cys	Phe
455	460		465
Leu Asn Asn Ile	Leu Val Ala Phe Asn	Asp Leu Arg Leu Cys	Cys
470	475		480
Pro Val Ala Leu	Ala Gln Asp Val Thr	Gly Ala Leu Glu Asp	Ala
485	490		495

Leu	Ala	Lys	Val	Thr	Lys	Ile	Ile	Leu	Ala	Phe	His	Arg	Ala	Glu
				500					505					510
Glu	Ala	Ala	Phe	Ser	Ser	Gly	Glu	Gln	Glu	Leu	Phe	Val	Gln	Phe
				515					520					525
Cys	Thr	Val	Phe	Leu	Glu	Asp	Leu	Val	Pro	Tyr	Leu	Asn	Arg	Cys
				530					535					540
Leu	Gln	Val	Leu	Phe	Pro	Pro	Ala	Gln	Ile	Ala	Gln	Thr	Leu	Gly
				545					550					555
Lys	Arg	Met	Lys	Ile	Leu									
				560										

<210> 21  
 <211> 219  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7503614CD1

<400> 21														
Met	Ala	Cys	Gly	Ala	Thr	Leu	Lys	Arg	Thr	Leu	Asp	Phe	Asp	Pro
1				5					10					15
Leu	Leu	Ser	Pro	Ala	Ser	Pro	Lys	Arg	Arg	Arg	Cys	Ala	Pro	Leu
				20					25					30
Ser	Ala	Pro	Thr	Ser	Ala	Ala	Ala	Ser	Pro	Leu	Ser	Ala	Ala	Ala
				35					40					45
Ala	Thr	Ala	Ala	Ser	Phe	Ser	Ala	Ala	Ala	Ala	Ser	Pro	Gln	Lys
				50					55					60
Tyr	Leu	Arg	Met	Glu	Pro	Ser	Pro	Phe	Gly	Asp	Val	Ser	Ser	Arg
				65					70					75
Leu	Thr	Thr	Glu	Gln	Ile	Leu	Tyr	Asn	Ile	Lys	Gln	Glu	Tyr	Lys
				80					85					90
Arg	Met	Gln	Lys	Arg	Arg	His	Leu	Glu	Thr	Ser	Phe	Gln	Gln	Thr
				95					100					105
Asp	Pro	Cys	Cys	Thr	Ser	Asp	Ala	Gln	Pro	His	Ala	Phe	Leu	Leu
				110					115					120
Ser	Gly	Pro	Ala	Ser	Pro	Gly	Thr	Ser	Ser	Ala	Ala	Ser	Ser	Pro
				125					130					135
Leu	Phe	Leu	Val	Glu	Leu	Leu	Gln	Glu	Val	Pro	Ile	Met	Thr	Cys
				140					145					150
Ser	Asn	Ala	Asn	Thr	Pro	Ser	Val	Asn	Thr	Gly	Tyr	Phe	Lys	Leu
				155					160					165
Ser	Ser	Val	Ala	Thr	Thr	Leu	Arg	Gln	Gln	Gln	Leu	Val	Leu	Glu
				170					175					180
Ile	Ser	Leu	Met	Ser	Val	Pro	Pro	Gly	Cys	Gly	Pro	Leu	Leu	Pro
				185					190					195
Val	Leu	Ile	Pro	Val	Ala	Ser	Phe	Cys	Cys	Ile	Ile	Thr	Ile	Trp
				200					205					210
Leu	Leu	Ile	Leu	Met	Phe	Glu	Lys	Asp						
				215										

<210> 22  
 <211> 497  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7503456CD1

<400> 22  
 Met Ala Ala Leu Thr Thr Val Val Val Ala Ala Ala Ala Thr Ala

1	5	10	15
Val Ala Gly Ala	Val Ala Gly Ala	Gly Ala Ala Thr Gly Thr	Gly
	20	25	30
Val Gly Ala Thr	Pro Ala Pro Gln Gln	Ser Asp Gly Cys Phe	Ser
	35	40	45
Thr Ser Gly Gly	Ile Arg Pro Phe His	Leu Gln Asn Trp Lys	Gln
	50	55	60
Lys Val Asn Gln	Thr Lys Lys Ala Glu	Phe Val Arg Thr Ala	Glu
	65	70	75
Lys Phe Lys Asn	Gln Val Ile Asn Met	Glu Lys Asp Lys His	Ser
	80	85	90
His Phe Tyr Asn	Gln Lys Ser Asp Phe	Arg Ile Glu His Ser	Met
	95	100	105
Leu Glu Glu Leu	Glu Asn Lys Leu Ile	His Ser Arg Lys Thr	Glu
	110	115	120
Arg Ala Lys Ile	Gln Gln Leu Ala	Lys Ile His Asn Asn	Val
	125	130	135
Lys Lys Leu Gln	His Gln Leu Lys Asp	Val Lys Pro Thr Pro	Asp
	140	145	150
Phe Val Glu Lys	Leu Arg Glu Met Met	Glu Glu Ile Glu Asn	Ala
	155	160	165
Ile Asn Thr Phe	Lys Glu Glu Gln Arg	Leu Ile Tyr Glu Glu	Leu
	170	175	180
Ile Lys Glu Glu	Lys Thr Thr Asn Asn	Glu Leu Ser Ala Ile	Ser
	185	190	195
Arg Lys Ile Asp	Thr Trp Ala Leu Gly	Asn Ser Glu Thr Glu	Lys
	200	205	210
Ala Phe Arg Ala	Ile Ser Ser Lys Val	Pro Val Asp Lys Val	Thr
	215	220	225
Pro Ser Thr Leu	Pro Glu Glu Val Leu	Asp Phe Glu Lys Phe	Leu
	230	235	240
Gln Gln Thr Gly	Gly Arg Gln Gly Ala	Trp Asp Asp Tyr Asp	His
	245	250	255
Gln Asn Phe Val	Lys Val Arg Asn Lys	His Lys Gly Lys Pro	Thr
	260	265	270
Phe Met Glu Glu	Val Leu Glu His Leu	Pro Gly Lys Thr Gln	Asp
	275	280	285
Glu Val Gln Gln	His Glu Lys Trp Tyr	Gln Lys Phe Leu Ala	Leu
	290	295	300
Glu Glu Arg Lys	Lys Glu Ser Ile Gln	Ile Trp Lys Thr Lys	Lys
	305	310	315
Gln Gln Lys Arg	Glu Glu Ile Phe Lys	Leu Lys Glu Lys Ala	Asp
	320	325	330
Asn Thr Pro Val	Leu Phe His Asn Lys	Gln Glu Asp Asn Gln	Lys
	335	340	345
Gln Lys Glu Glu	Gln Arg Lys Lys Gln	Lys Leu Ala Val Glu	Ala
	350	355	360
Trp Lys Lys Gln	Lys Ser Ile Glu Met	Ser Met Lys Cys Ala	Ser
	365	370	375
Gln Leu Lys Glu	Glu Glu Glu Lys Glu	Lys Lys His Gln Lys	Glu
	380	385	390
Arg Gln Arg Gln	Phe Lys Leu Lys Leu	Leu Leu Glu Ser Tyr	Thr
	395	400	405
Gln Gln Lys Lys	Glu Gln Glu Glu Phe	Leu Arg Leu Glu Lys	Glu
	410	415	420
Ile Arg Glu Lys	Ala Glu Lys Ala Glu	Lys Arg Lys Asn Ala	Ala
	425	430	435
Asp Glu Ile Ser	Arg Phe Gln Glu Arg	Val Glu Asn Asn Val	Ser
	440	445	450
Arg Asp Pro Ser	Arg Leu Tyr Lys Pro	Thr Lys Gly Trp Glu	Glu
	455	460	465
Arg Thr Lys Lys	Ile Gly Pro Thr Gly	Ser Gly Pro Leu Leu	His
	470	475	480

Ile Pro His Arg Ala Ile Pro Thr Trp Arg Gln Gly Ile Gln Arg  
 485 490 495  
 Arg Val

<210> 23  
 <211> 310  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7503459CD1

<400> 23  
 Met Cys Leu Phe Leu Pro Val Leu Ser Ser Glu Ser Ala Pro Leu  
 1 5 10 15  
 Val Val Arg Lys Gly Ser Asp Val Val Ala Gly Lys Met Ala Thr  
 20 25 30  
 Ala Ala Thr Ile Pro Ser Val Ala Thr Ala Thr Ala Ala Ala Leu  
 35 40 45  
 Gly Glu Val Glu Asp Glu Gly Leu Leu Ala Ser Leu Phe Arg Asp  
 50 55 60  
 Arg Phe Pro Glu Ala Gln Trp Arg Glu Arg Pro Asp Val Gly Arg  
 65 70 75  
 Tyr Leu Arg Glu Leu Ser Gly Ser Gly Leu Glu Arg Leu Arg Arg  
 80 85 90  
 Glu Pro Glu Arg Leu Ala Glu Glu Arg Ala Gln Leu Leu Gln Gln  
 95 100 105  
 Thr Arg Asp Leu Ala Phe Ala Asn Tyr Lys Thr Phe Ile Arg Gly  
 110 115 120  
 Ala Glu Cys Thr Glu Arg Ile His Arg Leu Phe Gly Asp Val Glu  
 125 130 135  
 Ala Ser Leu Gly Arg Leu Leu Asp Arg Leu Pro Ser Phe Gln Gln  
 140 145 150  
 Ser Cys Arg Asn Phe Val Lys Glu Ala Glu Glu Ile Ser Ser Asn  
 155 160 165  
 Arg Arg Met Asn Ser Leu Thr Leu Asn Arg His Thr Glu Ile Leu  
 170 175 180  
 Glu Ile Leu Glu Ile Pro Gln Leu Met Asp Thr Cys Val Arg Asn  
 185 190 195  
 Ser Tyr Tyr Glu Glu Ala Leu Glu Leu Ala Ala Tyr Val Arg Arg  
 200 205 210  
 Leu Glu Arg Lys Tyr Ser Ser Ile Pro Val Ile Gln Gly Ile Val  
 215 220 225  
 Asn Glu Val Arg Gln Ser Met Gln Leu Met Leu Ser Gln Leu Ile  
 230 235 240  
 Gln Gln Leu Arg Thr Asn Ile Gln Leu Pro Ala Cys Leu Arg Val  
 245 250 255  
 Ile Gly Tyr Leu Arg Arg Met Asp Val Phe Thr Glu Ala Glu Leu  
 260 265 270  
 Arg Val Lys Phe Leu Gln Ala Arg Asp Ala Trp Leu Arg Ser Ile  
 275 280 285  
 Leu Phe Ser Ser Gly Trp Pro Ser Ala Leu Ser Arg Lys Gln Phe  
 290 295 300  
 Arg Lys Gln Trp Arg Asn Ser Arg Lys Lys  
 305 310

<210> 24  
 <211> 1197  
 <212> DNA  
 <213> Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6024712CB1

&lt;400&gt; 24

```
gctaattggct ctcttggcgt tgcgacgtcc tggtcagcag tttttttcca ttctctccct 60
ccattttcttg agtgagcagc catgagttgg actgtgtctg ttgtgcaggc cagccggaga 120
gtgagctcgg caggagcgaa tttcctgtcc ctgtgtccca gtcaggcagc gcgcatgccg 180
ctcaagggcg cctggctctt ccccccggtg aagagtggagc ttgttgagcg cttcacttcc 240
gaggagcccg ctcatcacag taagggtctcc atcataggaa ctggatcggg gggcatggcc 300
tgcgctacca gcactcttatt aaaaggcttg agtgatgaac ttgcccttgt ggatcttgat 360
gaaggcaaac tgaagggtga gacaatggat cttcaacatg gcagcccttt catgaaaacg 420
ccaaatattg tttgtagcaa agattacctt gtcacagcaa actccagcct agtgattatc 480
acagaagggtg cacgtcaaga aaagggagaa acgcgcctta atttagtcca gcgaaatgtg 540
gccatcttca agttaatgat ttccggtatt gtccagtaca gccccctctg caagctgatt 600
attgtttcca atccagtggg taacttaact tatgtagcct ggaagttgag tgcattttcc 660
aaaaaccgta ttattggaag cggtgtaat ctggatactg ctcgttttcg tttcttgatt 720
ggacaaaagc ttggtatcca ttctgaaagc tgccatggat ggatcctcgg agagcatgga 780
gactcaagtg ttcctgtgtg gagtggagtg aacatagctg gtgtcccttt gaaggatctg 840
aactctgata taggaactga taaagatcct gagcaatgga aaaatgtcca caaagaagtg 900
actgcaactg cctatgagat tattaaaatg aaaggttata cttcttgggc cattggccta 960
tctgtggcgg atttaacaga aagtattttg aagaatctta ggagaataca tccagtttcc 1020
accataatta agggcctcta tggaaatagat gaagaagtat tcctcagtat tccttgatc 1080
ctgggagaga acggtattac caaccttata aagataaagc tgaccctga agaagaggcc 1140
catctgaaaa aaagtgcaaa aacactctgg gaaattcaga ataagcttaa gctttaa 1197
```

&lt;210&gt; 25

&lt;211&gt; 1001

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72176922CB1

&lt;400&gt; 25

```
gaaccagagt gtcagagcaa aaacctctct atctgcacat cctggggagc aaccgggcag 60
cgggagagct gcggccggcc cagtcccgct ccgcctttga agggtaaaac ccaaggcggg 120
gccttgggttc tggcagaagg gacgctatga ccgcagaatt cctctccctg ctttgcctcg 180
ggctgtgtct gggctacgaa gatgagaaaa agaattgagaa accgccaag ccctccctcc 240
acgcctggcc cagctcgggtg gttgaagccg agagcaatgt gaccctgaag tgtcaggctc 300
attcccagaa tgtgacattt gtgctgcgca aggtgaacga ctctgggtac aagcaggaac 360
agagctcggc agaaaaacgaa gctgaattcc cttcacgga cctgaagcct aaggatgctg 420
ggaggtactt ttgtgcctac aagacaacag cctcccatga gtggtcagaa agcagtgaac 480
acttgagctt ggtggtcaca gataaacacg atgaacttga agctccctca atgaaaacag 540
acaccagaac catctttgtc gccatcttca gctgcatttc catccttctc ctcttctctc 600
cagctttcat catctacaga tgcagccagc acggttcac atctgaggaa tccaccaaga 660
gaaccagcca ttccaaactt ccggagcagg aggtgcccga ggcagattta tccaatatgg 720
aaagggtatc tctctcgacg gcagaccccc aaggagtgaac ctatgctgag ctaagcacca 780
gcgcctgtc tgaggcagct tcagacacca ccaggagcc ccaggatct catgaatatg 840
cggcactgaa agtgtagcaa gaagacagcc ctggccacta aaggaggggg gatcgtgctg 900
gccaaaggta tcggaaatct ggagatgcag atactgtgtt tccttgcctc tcgtccatat 960
caataaaatt aagtttctca tcttaaaaaa aaaaaaaaaa a 1001
```

&lt;210&gt; 26

&lt;211&gt; 1174

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1392717CB1

```

<400> 26
gtggcacgca cctgtaatcc cagctactct ggaggctgag gcaggagaat tgcttgaacc 60
cgggaggttg aggttgagct gagccaagat cgtcccactg cactccagct tgggtgacaa 120
aacaagactc catctcaaaa gaaaaaaaaa acagcaccaa tgaagcctag ttctccacgg 180
gagtggggtg agcaggagca ctgcacatcg cccagtgga cctctggte tttgtctgca 240
gtggcattcc aaggctgggc cctggcaagg gcaccctgg ctgtctcttc atttgagac 300
cctgatcaga agtctctgca aacaaatttg ctcttgaat taagggggag atggcataat 360
aggaggtctg atgggtgcag gatgtgctgg acttacattg caaatagaag ccttgttgag 420
ggtgacatcc taaccaagtg tcccgatattg gaggtggcat ttctgacatg gctcttgggtg 480
taagcctgcc ttgccttggc tgggtgagtcc cataaatagt atgcactcag cctccggcca 540
caaacacaag gcctcgggga gggctagact gtctgcaaag gttttctgca tctgtaaaga 600
aaacaaggtg atcgaaaact gtggccatgt ggaaccctgg cttgtggggg actgtgtctc 660
catcttgact cagacagttc ctggaaacac cggggctctg tttttatatt ctttgatgtt 720
tttcttcttt agtagcttgg gctgcagcct ccactctcta gtcactgggg aggagtatatt 780
tttgttatgt ttggtttcat ttgctggcag agctggggct ttttgtgtga tccctcttgg 840
tgtgagtttt ctgacccaac cagcctctgg tttagcatcat ttgtacattt aaacctgtaa 900
atagttgtta caaagcaaag agattattta tttccatcca aagctctttt gaacaccccc 960
cccccttaat ccctcgttca ggacgatgag cttgctttcc ttcaacctgt ttgttttctt 1020
atttaagact atttattaat ggttggacca atgtactcac ggctgttgcg tcgagcagtc 1080
cttagtgaat attctgtata aatagacaaa atgaaaaggg tttgaccttg caataaaagg 1140
agacgttttg ttctggctct ttaaaaaaaaa aaaa 1174

```

```

<210> 27
<211> 948
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2701254CB1

```

```

<400> 27
aattggagtg gaattcccaa ggtgctggag gttgaaaggc attgagaatc taaaacgctt 60
tgcagacagc aagaccctct gtgtaagtta agaatgctca ctcagtccca gcagggtactc 120
agagggatcc ttttatttct ccagaacatt ctgcaggtga gctgggggaag ccactgggc 180
ctggcctcac ccccgagccc gagccttcag cctgggaatg gtctggcctc ttccttgctg 240
gtctctccagc ctggcctcgc aggaccctgg gcggggacccc aggaaccctc acccgctatg 300
tgcttcccca agaagcgtc cctgtggcct aatttgagaa aacaatgggc ctcaatccat 360
attaatgacc cttagaggac cctttgtcct cgggtgcacag gctgtaatca gcggggctcc 420
gggggctctg gcctaatttg gagggacagg ttttatcatc acccttgatt cgggtgaccc 480
aatctgacag gccacgacc ccctgtatgc ggggtccacg tcagagatgg gcttctctcc 540
agcggcccac ccagcgggc tggggagagg gaagggggag gtggtggcca tgggggaagt 600
tcgggggtgaa caaggggtca cagccaaacc ccacttgga tgggcctgtg atcgggttct 660
cgggaggagc aggatattga ttagatcagt gaatggtgtg gaggcagctc tccccaggca 720
cgtgctcccc accaccacc agcaagcgtc tggtgcctgc cggtgccagg gctgggtggg 780
gactctggga caggccggct gcctaggagg ggccaggcag gagccaatgg gctggctcca 840
gggactgcca ggagggcccc aggggaaggg agccgggaag cgttttggct ccttccggcg 900
tgaagttctg gaaacgtgtt tgcaggttag ggccgtggca tctctctt 948

```

```

<210> 28
<211> 2403
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 71774318CB1

```

```

<400> 28
cttttttttt ttttttaaaa gggaaaatga accatttttc agaatccttt agcccagtga 60
tcccatttct gggattttct tcccagaag gaaaatgtat tcaatatatg ggaaaggcta 120
tccgaaccaa gatgtcatt acagtgttat ttacaatagc ttaaaaaagg ataacaacct 180
aaaagatctc aaaatcagga aatggttaag taaatttcac aggggttaag gtgccatcat 240

```



```

gaggacactg cagtcacatg gaggaggcag gacactgatt ataaagctga gaaatgagga 300
caaggaccag gagagagtgt gcaaaagtga aaaacataat tgggtcagag ggtatggagg 360
tgagtatttt tctcaagttt ccaaactttc tgtaaatttg ctgtgttggc ttacagcat 420
agttggaaag aagtgtagcc cttecttccc atgctggggg agagagggtac aaaagcacca 480
ggtgagctct gcagctcttg gcttggaaac agtagaagga ggggcagggtg actgtgggag 540
ctgctctggc tggaccatgc tctgtgaccg catggcctgg atatcccagg gaagcctcta 600
acctgtagga gggagcccct catgctcttc ccagctggca ctctgagctt gtcacctcag 660
ccctatcgca ctctgtcct ggccagtttc tggttccctt gcttgggaca tcctgttcat 720
cctcaagtgg gcctgtgttt gtcccaagga caatcggtgc tctcgcttcc ccggactgcc 780
cagcatgcct cagcccaggc ctcaggccct tgctctctgt gctctggccc acgtgtcttg 840
cactgccatt cagaagcctg gtccctggaaa aaagggccca gctggcagcc ttttgagcaa 900
cctccctctc cctcccattt tctagaacct tccccactcc acaccctgga ttctgggtac 960
ctcacagctg ctgtactggg agaaacttgg cctgctgcca catttcttag atttgaaaag 1020
aaactttttg tttcctttta cattttaaag ttgtgaattt ggaaaaaaa aaaatctcat 1080
aatcgaagtg tttacccaag gaggcaatgt ttctttgttc cctaaaaggc tgttattaaa 1140
ttgatggatg gtgtgtctga caattgctgg catcttaaaa ctgaggaaat tcagcaattc 1200
cacactggag tgtctgtgga actaagctga ctgaggatga ccactagaga accttctatt 1260
tccttctaaa ggcaattatg tttctgtcct agaaatgggg ctgctccata gcaaacctgc 1320
ccaattctga gagcagagag aactgctgga gcatttgtag ccgacagtga aaggaagact 1380
ccacagacac tctcaagaac atctaaatga attcctagct ctatccataa aaaagcctag 1440
aaacagcaac ggaccaagta gcagtgaagc tccctagtgg ccagattatg gtctctaaat 1500
accattcccc actaaaagga accagggtt ctgggagaaa tggctgatgc cagggtctggg 1560
acaagaaata aaccagatga gcctgaagca tctgtcacat cagaaggcaa gaaaggtgac 1620
cgtttaaaag gattcaggac ttcccagcat ccaagaggga atgatttaaa catcaatcag 1680
gataaacact gcaaagaatt gactcttcaa atgcagtact gtatgtttaa aatccatgag 1740
ttcataataa tccctaaaaa gaaacacacg ttttagacga tgctagagaa acagtatttt 1800
ggaaagccaa aaagaaaagg aggggagaga atccaacact taggtgcct tccccatctg 1860
aatttcacca cagagtagcc aaataattga gggaaattct ctctttattg cagtattttc 1920
agcaataaaa tttaaaaaag aatagtaggt gaaaatacca ccatttttca acccctaattg 1980
aattaatgga ctctgccagc atgactggct tccaacatca caaagagcct ggcaaccaga 2040
cattctatct tgtatgtccc cagggccatc ttgccctagt agtcaaacct caatctgatc 2100
aggcctctag atctaattaa caatgtacta gaaatacagg ggagcaagga gcagactaaa 2160
caataccata gagatgcagc cagtaaaatt cagggtccacc tgggtttcttc aacaggtaat 2220
ttacaagaaa aaaatggaaa aaggggaaag agaacctaga gaggtttaag agacatagca 2280
accaatcaca acacatagac cttatctgga tctcaaatac aatgagcaaa ctgttaaaaa 2340
gggtgggtgt ggggactatg aatctgctag atatctggta atttaagaaa tcattgttaa 2400
ctg 2403

```

&lt;210&gt; 29

&lt;211&gt; 2848

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 71802522CB1

&lt;400&gt; 29

```

gcctggatct caacaatacc aggcattctg tagaactgtg tgattcctcg ggggaataaa 60
agaaaaatga acactagaaa attcatttat agacattcag gccaaattta gaaaacgtta 120
caacatttcc tcttgggaca aaacagtttt taacatttat ttaagcattc cgtttataat 180
atgagtagaa aaagagcgcg attaacttag taagaaaaga tctatgcttt ttacagcttt 240
ctgttttgat ataatacatt aatttatcat gttttatagt ttcagctgtg ccacagtcca 300
aagtttctga agccatgac gagccaaaat tttatgccca gcaaacacag aaattaatag 360
tacaccttgt ttatttttgg ttaaaattat cctaatttcg gattaaaaat aagattgtta 420
atatttttaa tcaaaacctt tctcctccgc aattccagct ttacgagatc agaaagtggg 480
gcaacagcgc tctctagtgg ccaaatgcga cctccagtc actgaattta atttttatc 540
atctcaaacc aattctaate tcattcttca atttgtggct tggactgctg acactaataa 600
tgcttcatc ctgtgactgt caacaagtca agagcagcaa gcaggagctc ctgcatgcct 660
aggatattct aatcagtgt cccaacagct tccagtagtg tcttttccaa aagcttccag 720
cccttctttt gccagacacc aaaagtcccc atgctgtaac cgcccagctg tgccttttaa 780
aacaattata ggacaggcta gatttaacaa aacctacaac tttctaaata tcaactgcc 840
agaggctttt tattttccag aggaaccaga cactttggca accaagaaat tacagaaatc 900

```

```

aagtattgtt atgtttttgt ttttctctta cagctgttac ttttaggatt aaatcatgaa 960
tatttgctac ttttaacactg taactttaat gctttggggc ccacttaggt ggccgatttt 1020
ggggggaatc gaagccaaat caaaagcttt ttgctggatc tctgagtggc tgtctggaaa 1080
gtccctttgc gcttctgttt gttggaattt ctatggacaa taaatgggat tttaaaaata 1140
tatatcaaaa gtcatgaac taagcattcc actagaattt tttttaaag caatgactat 1200
tcaattggct taagtgtttc ctgttaggtt aatcttttta ctcatcttca aaactgcttt 1260
cctccagttt aggtgtaaca attgtatttt tttataagt aatagacata aatttcaaca 1320
acattttgct tcaaagtttg gagtggaggg ccagagattc ttgtgaatag gcatctaact 1380
tcatcttttc ccgcgaacca cacttctggg agactctcac ctgggtaaat gacctaagat 1440
cactaaccac cctatcccca gataacaggc caattttttt gtttgtttgt ttctaaagaa 1500
ccagctaaac cgtggtcata gtcaaaatat tgtaaaatgc catctcctta ggctacttcc 1560
attacctggg tcatttcaaa aatcagaaat tcattcacct ctctcttctt acctctttca 1620
tgtttacttc ctagggtgctc acacctcacc tgggggctcc catgcccctg ccttcttggg 1680
ccaacgcagt ctgtgtttata ttttattgtc tagattttta gaccagttt tgtctgcatt 1740
tttctttttc tcccacaaac acaaacatca aaatgagatg gtttcaaagc aaagcactgg 1800
gcctgtcgta aactcattac ctccagacgt ctggcccgct agatggcttt atgacacctc 1860
ggctcttctt gttttcaaag agcccttaggt accgagctgg ccacagtggc aggggtgcgc 1920
agcaccttct gccggatctg gcccttctct ggctttccct ccagaccccc ctatgctttt 1980
tcttcgcgct ccctctttca ttgggatccc ccaagatctc tgccaccgcc ccaccttcc 2040
accagctca gccacttgg cagtgtctgcc tctttggcct gcaaatgctg tgctcaccca 2100
aaccctcact tacaatgacc ttcattttag ccctgaatg cagccctcag agggccaagc 2160
tgggggctaa acatacacag aagctgggtg gggggaaggg ggcggtgaaa tggaggtggc 2220
tggggaggag agccttgacc attttgatcg ctaagggtcac tttggggctt tgggtggggag 2280
gagcagaagc tcacagcctg accagctggg acctcccga gccagctagc ccacccgagc 2340
tgggacagtt gctccagagt gtagagttgg cctttccctt gtttggagag gggtttggca 2400
tctggggctt caggagcccc gggaagggtt ggggtgctgtg taccgaagc ccagcctgaa 2460
agtgcctggc ttctggggcc tgctggctct tctggcgct tctgcaaggc taacactggg 2520
cccaagcaca gaggctcgtg aggacctga accaggcttg gggaggcacc aatgtcccc 2580
cacagggtct gtgtgctgc atctgcctac acaaaggat gtgaggcaga gaaggcagag 2640
acacatcaga gagatctgtc aggtggaagg gcctgaatt ttaacaatg tctactccc 2700
aggggtagag gagaaatata ctgagtgtga tattttgcag agaggcagag aaagcaggca 2760
cacaatgttt agggccagcg tctgggctcc atttgatcag gtcagcattt attagtagga 2820
agcagtaaca tttacaactg gtccctcgg 2848

```

&lt;210&gt; 30

&lt;211&gt; 3394

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6425956CB1

&lt;400&gt; 30

```

ccgggggaag gacactggag gccggtgggc catagttgta ccgggcccac gatgtccgtt 60
tcaggatcca ccgcaataac gtgtctgaca agaatgccg ctgacagtgc ttctgaagct 120
gcgtgggact tgctgggtcaa agcacggcag gctctgtgct ggcaaaccag ggccccgcgg 180
cagccccaag gccagcctgc agacgccagt gtgccggct tcttcccatg ttctgagatg 240
ggggagggga aggtgtaaaa gagagctgga tgggggctgt ggggactgag ggtcctggga 300
ggctggccag gggctcagca agcggccact ttgggaggcc ttgccctctg ggctgtgggt 360
ggacagcact aggcctgcag gctgggctgc caggaggaag agcagctccg gacctcaggc 420
ccccccagca gccggcgatg gaggggtgat ggagaattga cgagtgtctg cgggggtctca 480
ctcgggaaaag gcccggttaa ctctgagcaa ctgggagcct cgctcccacc ccagtgcctg 540
gaaacacaaa cagtatttta ttgagccttg gctgccgggg ctctggtttc aggaaaagcc 600
tcagtctctg atttgtgaat ttttaaggag gcaaacaagg tgaatttgac ctctctagga 660
gtgggagtta gggaaaattc aggtgcccct accaagccct tggaaaccgtg cagcaggggt 720
caggggaggg atgggcaaaag gcggtttggc ccatggagca ggtctccttg tcttcccgga 780
gcacgggggg cggggagctc cagctcttca tcaagcgccc tttgggggtt ccaactgttt 840
tctgtgttct ctacatcact cgggtctcat gtctgaaggg cctgtctcac ctgccacctc 960
cctggccttg gcctcgacat ctgagcgctt catcacctcc tctccacatg ccaggggggt 1020
ccctcccaa ggctggctgg gcaggctcca cggactaggc ccaggaggga gcagtgggct 1080
tccccagga aagagcagag ccagcactgc ctgcctaggg agggcaccca ctaccaggca 1140

```

tggtctggtgg	ctccggttaa	agaaatcttt	gtccatgtgg	gagtgggagg	tgttctctca	1200
tcctgcctgg	aagccacgtc	ctggctctta	ccgtggactc	tgcaacagcc	gtggtggcca	1260
catgaagatg	gaggaacctg	gtggctcagg	ggccctgat	gtcaccgcct	ccaaggcaac	1320
gggactgggc	agggcagctc	ctcaggaagg	atgatccac	ctgcacgggt	cagccctgca	1380
gggggacaag	gccccaccac	acccacctc	aggccacca	ctggacctgt	cctgccaaga	1440
aggtcccttc	caatttgtcc	ggcagctgg	tccttgata	cagcccagac	gcagcatcat	1500
cctcatgttc	acctctgtcg	tgccctactt	ccccgtaaa	tgacaactgt	gtagtcaccc	1560
acgccgggca	agagaacata	ggtgccacct	ggcggggccac	acttttcagg	ggccgtgcct	1620
cgggcaggcc	cggctctctg	ttgtcccgca	gcctcagg	ctggcggtgg	ctgctcagt	1680
ccataggagt	gcccttgcca	cacagctggg	cagcgagcat	gactcagcgt	ttacactcaa	1740
ccagccgcac	aggtccaga	ggggaagtcc	tctgggtcc	ttctatgaga	gagcatcctg	1800
tctgcagccc	tggtctggct	ccctccacat	cattcatgca	ttcatccac	gaggactatg	1860
gtaccgtggg	ctcgctctag	ggaagcaagg	caggagcgca	ctggggctac	ttggattccc	1920
ggggcagaca	catatgtacc	ctgcacaccc	tcaaatgagc	aaagccaccc	cactggcccc	1980
tggtgaggc	catgtggacc	cccacacctt	tggggccggc	tgggcaaac	ttgaaccccc	2040
aatttctgcg	tggcctttgg	ctgcctctct	tctccaaggc	gtgactctta	ctccagagac	2100
tcaggcgagc	acgtgtccct	taccttattt	tctctcaatc	aaactgaaac	ccatttgtgat	2160
cccccatagt	ccagtgcggg	ctctgttatt	attacgggtg	tctcccttcc	tcccgtgcc	2220
caggacaggc	catgagccag	agatacaagg	ggcccaggc	aagatgcagg	gctctctgct	2280
cctggctctt	tatcgctcgt	gggacaccct	ttgtccaaac	tcaaggaatc	ccgggagggtc	2340
tggctttgcc	gctttggctg	ggactcagg	acctgggcat	acctgcatct	gtgccccctc	2400
ttcgtcccag	tgagacggac	cctcagtatc	ccccccaaca	tagaacagga	aggtctcaag	2460
ccagccccct	cctccagttc	agtcaccctc	cctcacatcc	accagagag	gtccagacct	2520
gtggccactt	ccatgccctc	tccagaccag	gagacacctg	ctgctgacct	cgtggaaaaac	2580
ttagattttg	acattctgat	gcttcggaag	tggtggctcc	tcctccctca	ccctccgcc	2640
acctgtgggc	ctcctctctg	catctcaaga	gaacaaccag	atctttgggc	tcctgggggtg	2700
tgtgccatgc	aatttagacg	aagtgtcttg	aaaatatgcc	attcagtctc	tgactaggaa	2760
aataagtctg	acctgatagg	tctgatgtca	tcagctcttc	aacatgagac	aaaagagggg	2820
attttatgtt	ttgagtcatt	agaatgatat	aataattttc	tgaattgaca	tctggatgtt	2880
gaaattagga	tgggtgcaaaa	ggggtccagg	gcctcaggct	gggcgcagca	gccagctccc	2940
aatgacgcag	aagctgcttc	aaaaccccc	caacaaagag	gggcacatgc	aagtcacca	3000
agtgggaagc	cttcaccaag	gccacacca	aagctcactg	attgtctgtc	caaagtctct	3060
tgattcctgg	ccatgaacaa	gcacaataga	aaaagacaca	gggtcctagt	ggctacaagt	3120
caatgtgaat	tggcacatgg	tctagcagtt	ttaaaatctg	acagtagagt	atggcaatgg	3180
gcaagggccca	agaagtccctg	agatgggagg	tcagcgctct	aactgggctc	agtgagggtc	3240
tgtgaccagt	gtctggacac	tagctacagg	ggaccgggca	gaggattctg	ggcagaagga	3300
aaatgtctaa	aagtactttg	ggaggggtaaa	ggacaggggc	cttattttaa	ataaagactg	3360
aaggataata	agtgtaaaaa	aaaaaaaaaa	aaaa			3394

&lt;210&gt; 31

&lt;211&gt; 1858

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7494288CB1

&lt;400&gt; 31

gctgcagggt	gggtcaggga	gggtcgtgg	gtaacagcaa	tgtgagcctg	gggtctgaac	60
ctaccatggc	ccacacctga	cagacagtaa	gcagagaact	ttgtggttgg	gggagcaaa	120
gaaagagaca	aagataagcc	aagcgtccta	ggtgaaggga	caaggaggtg	tcaagggcac	180
aggattttct	tgcttgagtg	ctttccttgt	tgacagaggg	aatttggcct	ccacctgctg	240
aacacagAAC	tgcgctgagg	aaggcactgg	cttgagagaag	atgccagagc	tctgccagct	300
gacactcagc	ctgcaccttc	aggactcatg	acttcttctg	gcccagagg	cccagacagt	360
gggcatcacc	tggggctctc	ctctcttttg	gaaaaaaaga	ggcttcctgg	ggattccagc	420
agccactgag	gctgaggact	ctcaggcttt	taggaggccc	atggagactc	acttctctcc	480
gggccccttt	cccaaaacac	ccaccctggg	acccccttgg	actccagccc	accaggccac	540
caggacaaat	ggcctctgag	accttcaaca	ctgaagaccc	aggtgcagcc	ccggcttctc	600
aatcagtcct	agacgtggga	caggacttaa	gttttctcag	aagaagtgcc	aaggatcttt	660
aaagggcacg	ttcatctctc	ttcagcccct	ccctccagga	cacagcacca	agaggaaaca	720
tccagcaggc	atgggtgaga	gagacacaaa	agatctgcag	gaactcttcc	aggacttgct	780
ggcagctgac	ctagccccat	tcttgctccc	tccctttctc	aggttttgat	catcattcca	840

```

gaccctaggg gacctcaact ggggtgtcttg ccccttaggc tccggaagg gacctcccg 900
ggatctcagg aagccctctg gtgctcagag gctcctggga aagtccttag ccatgatacc 960
acatgctcag aagccccacc ttcacagacg ctggccccag atgtagctgc ctccctgtgt 1020
cacagactct cgattccatg gacacagtc tcatgggctc cctccagcac tgctgttgcc 1080
tgctgcctaa gatgggtgac acttggggcc agcttccctg gcccgggcca cccacccag 1140
caatgctgct gatctccctc ctcttggcag ccgggttgat gcaactcgat gccggacca 1200
gctgccccgt cctttgcaca tgccgtaacc aggtgggtgga ttgtagcagc cagcggtat 1260
tctccgtgcc cccagacctg ccaatggaca cccgaaacct cagcctggcc cacaaccga 1320
tcacagcagt gccgcctggc tacctcacat gctacatgga gctccagggt ctggatttgc 1380
acaacaactc cttaatggag ctgccccggg gcctcttctt ccatgccaag cgcttggcac 1440
acttggacct gagctacaac aatttcagcc atgtgccagc cgacatgttc caggaggccc 1500
atgggctagt ccacatcgac ctgagccaca acccctgggt gcggagggtg catccccagg 1560
cctttcaggg cctcatgcag ctccgagacc tggacctcag ttatgggggc ctggccttcc 1620
tcagcctgga ggctcttgag ggcttaccgg ggctgggtgac cctgcagatc ggtggcaatc 1680
cctgggtgtg tggtgcacc atggaacccc tgctgaagtg gctgcgaaac cggatccagc 1740
gctgtacagc aggtaataga ggggcagaac ggggcagtca acaggagggt cttgcctcaa 1800
tgggaagcaa agtctccaaa gaaagcgggg gaacttagag ccagaaagca cctactgc 1858

```

<210> 32  
 <211> 1242  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474330CB1

```

<400> 32
aatatagcac gaccctgtgt ccaacaacaa caacaacgac aacacaaaca aacagaaaac 60
taaagacatg cttattgacg ttccaatttc ccaatctcag cccaataga cactttaaag 120
ctgttactca gggcagtgta ttcgggggtga tgaggccatg cccgcttctt ctcttccagg 180
aaagctctgg ttcgtctca cgatgctgct gcggatgctg gtgattgtct tggcgggcg 240
accgctctac caggacgagc aggagagggt tgtctgcaac acgctgcagc cgggatgcgc 300
caatgtttgc tacgacgtct tctccccctg gtctcacctg cggttctggc tgatccaggg 360
cgtgtgcgtc ctctctccct ccgctgtctt cagcgtctat gtctgcacc gaggagccac 420
gctcgccggc ctggggcccc gccgctgccc cgacccccgg gagccggcct ccgggcagag 480
acgttgcccc gggccattcg gggagcgcgg cggcctccag gtgccgact tttcgggcgg 540
ctacatcatc cacctcctcc tccggaccct gctggaggca gccttcgggg ccttgacta 600
ctttctcttt ggattcctgg cccgaagaa gtctccttgc acgcccctc cgtgcacggg 660
cgtggtggac tgctacgtgt cgcggccccc agagaagtcc ctgctgatgc tgttctctct 720
ggcggtcagc gcgtgtctt ttctgctggg cctcgccgac ctggtctgca gcctgcggcg 780
gcggatgcgc aggagggcgg gacccccac aagccctcc atccggaagc agagcggagc 840
ctcaggccac gcggaggagc gccggactga cgaggagggt gggcgggagg aagagggggc 900
accggcgcgc ccgggtgcac gcgccggagg ggagggggct ggcagcccca ggcgtacatc 960
cagggtgtca gggcacacga agattccgga tgaggatgag agtgagggtg catcctccgc 1020
cagcgaaaag ctgggcagac agccccgggg caggccccac cgagaggccg cccaggaccc 1080
caggggctca ggatccgagg agcagccctc agcagccccc agccgcctgg ccgcgcccc 1140
ttctgcagc agcctgcagc cccctgaccc gcctgccagc tccagtgggt ctccccacct 1200
gagagccagg aagtctgagt ggggtgtgaaa aaaacagcac ct 1242

```

<210> 33  
 <211> 544  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5911370CB1

```

<400> 33
gattgatcgg tcacctgacc tcatagataa ctctagggcg gacggacgga cgcgctctc 60
cggctcccgtc cgtaatagca ctgatccgat ccacgccggc cggcgatgga gctcatcaag 120
tccagggcga ccgtgtgcgc gctcctcctg gcgtgctcc tgctctcgca ctacgacggc 180

```

gggacgacga	cgacgatggt	ggcggaggcc	cgggtgtgca	tgggcaagag	ccagcaccac	240
tcgttcccc	gcatctccga	ccgcctctgc	agcaacgagt	gcgtcaagga	ggacggcggg	300
tggaccgcgc	gctactgcca	cctccgctac	tgcagggtgcc	agaaggcggtg	ctaagcaaaag	360
ctcttgaaac	acccttggct	tgccagaact	gaactgtggt	agtactaagt	aacacccttg	420
gctagctgtg	cacaacctac	gtaccgtgca	tgcattgta	gtgggtgtcat	gtaacgtgac	480
agcaataaat	attaataaca	ataataacac	ggcatgtagc	ctttgcatgc	ttctaaaaaa	540
aaaa						544

&lt;210&gt; 34

&lt;211&gt; 3471

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7647134CB1

&lt;400&gt; 34

cgggggcctg	gccgcgcgct	ccccccccgc	aggcgccacc	tgggacatcc	ccgggattgc	60
tactttctctg	ccaacttcgc	caactcgcca	gcacttggag	aggccccggt	ccccccccgc	120
cgccctctga	ccgccccgcg	cccgcggcgc	tctccgacca	ccgcctctcg	gatgaccagg	180
ttccaggggg	gctgagcgag	tcgcctcccc	cgcccgagctt	cagccctggc	tgcagctgca	240
gcgcgagcca	tgcgccccca	gtgcaccccg	gcccaccgcc	ccggggccat	tctgctgacc	300
gcccagcccc	gagccccgac	agtggcaagt	tgcggctact	gcagttgcaa	gctccggcca	360
acccggagga	gccccacggg	gaaggcagtc	gtgcgcccc	cgcccccggt	cgcccccgag	420
cagccggggc	ttcactcctc	ctccctcccc	caccgtccct	cccttttctc	ctcaagtcc	480
gaagttgagt	ttgagaggcg	acacggcggc	ggcgccgcgc	tgctcccgct	cctctgcctc	540
cccatggata	tgcactgcaa	agcagacccc	ttctccgcga	tgcacccagg	gcacgggggt	600
gtgaaccagc	tcgggggggt	gtttgtgaac	ggcgggcccc	tacccgacgt	ggtgaggcag	660
cgcatcgtgg	agctggccca	ccagggtgtg	cgccctctgt	acatctcccc	gcagctgcgg	720
gtcagccacg	gctgtgtcag	caaaatcctg	ggcaggtact	acgagaccgg	cagcatcaag	780
ccgggtgtga	tcggtggctc	caagcccaaa	gtggcgacgc	ccaaagtgg	ggacaagatt	840
gctgaataca	aacgacagaa	cccgaactatg	ttcgccctggg	agattcgaga	ccggctcctg	900
gccgagggca	tctgtgacaa	tgcacacagt	cccagcgtct	cttccatcaa	cagaatcctc	960
cggaacaaag	ttcagcagcc	tttccaccca	acgcgcggtg	gggctggggac	aggagtgcac	1020
gccccctggc	acaccattgt	tcccagcagc	gcctccccct	ctgtttccag	cgccctccat	1080
gaccagtggt	gacccctact	catcaatggg	atcctggggg	ttcctcgctc	caatgggtgag	1140
aagaggaaac	gtgatgaagt	tgaggtatac	actgacccct	cccacattag	aggaggtgga	1200
ggtttgcctc	tggtctggac	tttaagagat	gtgtctgagg	gctcagtcct	caatggagat	1260
tcccagagtg	gtgtggacag	tttgccggaag	cacttgccag	ctgacacctt	cacccagcag	1320
cagctggaag	ctttggatcg	ggtcttttag	cgctcttcc	accctgacgt	cttccaggca	1380
tcagagcaca	tcaaatacga	acaggggaac	gagtaactcc	tcccagccct	gacccctggg	1440
cttgatgaag	tcaagtcgag	tctatctgca	tccaccaacc	ctgagctggg	cagcaacgtg	1500
tcaggcacac	agacataccc	cggtgtgact	ggtcgtgaca	tggcgagcac	cactctgcct	1560
ggttaccccc	ctcacgtgcc	ccccactggc	cagggaagct	acccacacct	caccctggca	1620
ggaatgggtg	ctggggagcg	gttctccggc	aacccgtaca	gccaccccc	gtacacggcc	1680
tacaacgagg	cttgagagatt	cagcaacccc	gccttactaa	tgcgcgcccc	cggtccgccc	1740
ctgccgctcg	tgccgctgcc	tatgaccgcc	actagttacc	gcgggggacca	catcaagctt	1800
caggccgaca	gcttcggcct	ccacatcgtc	cccgtctgac	ccccaccccg	aggagggagg	1860
accgacgcga	cgcatgcctc	ccggccaccg	ccccagcctc	accccatccc	acgacccccg	1920
caacccctca	catcaccccc	ctcgaaggtc	ggacaggacg	ggtggagccg	cgggggcgga	1980
ccctcaggcc	cgggccacc	gccccagcgc	ccgcctgccc	ccccctcccc	cctgcctgga	2040
ctgcgcggcg	ccgtgagggg	gattcggccc	agctcgctcc	ggcctccacc	aagccagccc	2100
cgaagcccg	cagccaccct	gcccgtactcg	ggcgcgacct	gctgggtgcg	gccggatggt	2160
tctgtgacac	acaatcagcg	cggaccgcag	cgcgccccag	ccccgggcac	ccgcctcgga	2220
cgctcggggc	ccaggagctt	cgctggaggg	gctggggcaa	ggagattaag	aagaaaacga	2280
ctttctgcag	gaggaagagc	ccgtgcccga	atccctggga	aaaattcttt	tccccagtg	2340
ccagccggac	tgccctcgcc	ttccgggtgt	gccctgtccc	agaagatgga	atgggggtgt	2400
gggggtcccg	ctctaggaac	gggctttggg	ggcgtcaggt	ctttccaagg	ttggggacca	2460
aggatcgggg	ggcccgagcag	cccgcaccga	tcgagccgga	ctctcggtct	ttcactgctc	2520
ctcctggcct	gcctagttcc	ccagggcccg	gcacctcctg	ctgcgagacc	cggtctctcag	2580
ccctgccttg	cccctacctc	agcgtctctt	ccacctgctg	gcctcccggt	ttccccctct	2640
gccagtcctt	cgctgtccc	ttgacgcctt	gcactctctt	ccctgactcg	cagccccatc	2700

```

ggacgctctc cgggaccgc cgcaggacca gtttccatag actgaggact ggggtcttcc 2760
tccagcagtt acttgatgcc ccctcccccg acacagactc tcaatctgcc ggtggtaga 2820
accgggtctc agctggcgtc tgagctgctg cggggtggaa gtggggggct gccactcca 2880
ctcctcccat cccctcccag cctcctcctc cggcaggaac tgaacagaac cacaaaaagt 2940
ctacatttat ttaatatgat ggtcctttgca aaaaggaaca aaacaacaca aaagcccacc 3000
aggctgctgc tttgtggaaa gacgggtgtg gtcgtgtgaa ggcgaaaccc ggtgtacata 3060
acccctcccc ctccgccccg ccccgccccg ccccgtagag tccctgtcgc ccgcccggcc 3120
tgctgtaga tacgccccgc tgtctgtgct gtgagagtcg ccgctcgctg ggggggaagg 3180
gggggacaca gctacacgcc cattaagca cagcacgtcc tgggggaggg gggcattttt 3240
tatgttacaa aaaaaaatta cgaagaaaga atctcatttg caaaatagcg aacatggtct 3300
gtgactcctc tggcctgttt gttggctctt tctctgtaat tccgtgtttt cgctttttcc 3360
tccctgcccc tctctcctc tgcctctc tctctcgc ttctctcccc ctctgtctct 3420
gtctctctcc gtctctgtcg ctctgtctg tctgtctctg ctctttctcg c 3471

```

&lt;210&gt; 35

&lt;211&gt; 1484

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1631327CB1

&lt;400&gt; 35

```

gtggggtgca aggagccgag gcgagatggg cgtcctgggc cgggtcctgc tgtggctgca 60
gctctgcgca ctgacccagg cgggtctccaa actctgggtc cccaacacgg acttcgacgt 120
cgcagccaac tggagccaga accggacccc gtgcgcgggc ggcgcggttg agttcccggc 180
ggacaagatg gtgtcagtc tggtgcaaga aggtcacgcc gtctcagaca tgctcctgcc 240
gctggatggg gaactcgtcc tggcttcagg agccggattc ggcgtctcag acgtgggctc 300
gcacctggac tgtggcgcg gccaacctgc cgtcttcgc gactctgacc gcttctcctg 360
gcatgacccg cacctgtggc gctctgggga cgaggcacct ggctcttct tegtggacgc 420
cgagcgcgtg ccctgcggcc acgacgacgt ctcttttccg cctagtgcct ccttcgcgt 480
ggggctcggc cctggcgcta gccccgtgcg tgtccgcagc atctcggtc tgggcccggc 540
gttcacgcgc gacgaggacc tggctgtttt cctggcggtc cgcgcgggce gcctacgctt 600
ccacgggccc ggcgcgctga gctgtggccc cgaggactgc gcggacccgt cgggctgcgt 660
ctgcccgaac gcggaggcgc agccgtggat ctgcgggccc ctgctccagc ccctgggccc 720
ccgctgcccc caggccgctt gccacagcgc cctccggccc caggggcagt gctgtgacct 780
ctgtggagcc gttgtgttgc tgacccacgg ccccgcatth gacctggagc ggtaccgggc 840
gcggatactg gacaccttcc tgggtctgcc tcagtaccac gggctgcagg tggcctgtgc 900
caaggtgcca cgtcgtcccc ggctccgtga ggcgatacag gagatccagg tgggtgtggt 960
ggagaatggg cccgagacag gcggagcggg gcggctggcc cgggcccctc tggcggacgt 1020
cgcgagaac ggcgagggcc tcggcgctct ggaggcgacc atgcgggagt cgggcgcaca 1080
cgtctggggc agctccgcgg ctgggctggc gggcgcgctg gcggctgccc tgcgtctggc 1140
gctgctggtc ctgctgggtg cgcgcgcgct gctgcgcgc gcggggaggc tcagggtggg 1200
gagggcacgag gcggcgggcc cggctggagc gccctcggc ttccgcaacc cgggtgttca 1260
cgtgacggcc tccgaggagc tgccccctgc gcggcggtc agcctgggtc cgaaggcggc 1320
cgcagacagc accagccaca gttacttcgt caacctctg ttccggggg cagaggccga 1380
ggcctgagcg gccgcctgac cgtcgacctt ggggctctcc acccctctg gccccagtcg 1440
aactgggggg ctagccacct cctcgtccag cccccaacc tccc 1484

```

&lt;210&gt; 36

&lt;211&gt; 1773

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 044232CB1

&lt;400&gt; 36

```

ctcagggcct ggcagggtccc agaagggtggc gagtttccgc gccagaggct tacagggtcca 60
ggtggagagg ccgggctggc cagggtctcg gcctccggcg tcgggaaatg gcggcggggg 120
gcaggatgga ggacggttcc ttggatatca cccagagtat tgaagacgac ccacttctgg 180

```

```

atgccagct tctccacac cactcattac aagctcactt tagacccoga ttccatctc 240
ttctacagt catcatagtg aatcttctgt ggtttattca tctcgtgttt gttgttttag 300
catttttaac aggtgtgctt tgttcttate ctaatccaaa tgaggacaag tgcccaggaa 360
attacacaaa cccattgaaa gttcagacgg ttataatcct tgggaaagtt attttgtgga 420
ttctccattt actccttgaa tgctacatcc agtatcacca cagcaaaatc agaaaccgag 480
gctataactt gatctaccga tcaacaaggc atctcaagag acttgcggtg atgatacagt 540
cctctggcaa cacagtgtt ctcctcatalc tgtgcatgca gcactccttc ccagagcctg 600
gcagattgta tcttgacctc attctggcca tcttggcact ggaactcatc tgttccctga 660
tatgtctcct catttacaca gtgaaaatcc ggagatttaa taaagctaaa ccagagcctg 720
atatacttga agaagaaaaa atctatgtct accccagcaa tattacctcg gagactggat 780
tcagaactat ttcaagccta gaagaaattg ttgaaaagca aggagacacc attgaatacc 840
tgaagcgaca caatgcgtg ctgagtaagc gattgttggc tctcacttcc tcagacctgg 900
gctgtcagcc aagtagaacg tgagaggctc acggtcatga cagcaattgc agaggaaccc 960
agagtaattg agactgactg accacctgac aagctgccac ggggaactgc agcttttgct 1020
gaatagcatt ttacagtgtt tgttggaaac ctgaatttgg ttctgacttc tgtggctgtt 1080
taaaatatag ggcttgtggg tcaattgaaa agtacctgta gaagcccgga taacttgagg 1140
ggaatgtctg tttgtacttt tggaaattat ttaactgctt ggtttatcct aggatcagag 1200
gctaaacaac tccatagtca acacttttcc acctgacatt agaaccctgt aatgatttca 1260
ttatggatag gggaagtcta acaagacaat cgattttaaa caggttttta atcaaataag 1320
ttcttcccac ttttaagctg atgaaagaag tcattatttc ttgtgaatat tttttcctgg 1380
agagccttat catgtatttt atatgcttat gtggtgttgg atgacatcat ggaccatata 1440
gctttttatag agaatttttc tcaccatagg actgaggctc caccagggtga tctactatgc 1500
aaattcctac agttttctat tcttaagaaa taagggtctg gcacggcgga tcatgaggtc 1560
aggaaattga gaccatcctg gctagcacgg tgaaaccctg tctctactaa aaatacaaaa 1620
aaaattagcc gagcatgggt gcgggcacct gtagtcccag ccacctggga ggctgaggca 1680
ggagaattgc ttgagcctgg gaggcggaga ttgcagttag ccgagatggt gtcactgcgc 1740
tccatgcctg gcgacagagc agagcgagac tca 1773

```

<210> 37

<211> 2016

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 560293CB1

<220>

<221> unsure

<222> 1719

<223> a, t, c, g, or other

<400> 37

```

atagaattcg gcacgagggg agagttctac gagggagggg aagcggttgg acgtgttcgc 60
ttgggttcct gctgcggcag ctacctcgca atctctctgc atcgatcgcc gctcgcaagc 120
tactgaccgt actcgggcgt attaggagcc gcgttccagc ctacaccccc acggtgctgt 180
tttcgacttc agaaaggatc tagtctcagc acagaagcgc ctcaggcgcg gcgcaaagct 240
cgagcggaag gcggggggcg ccggagcctc tctcggggga gccgcgcctg aggaggcgga 300
agaaccccc tgacgcgact ggcgtgtgct tctgcccgcc accgccccct ccgctctcac 360
ccggggccgt cctggccact gcccctgcgc cggaggcagc ggccggcagc gctctccttt 420
ccacagccgg cgctccgcga cccgcttggc tcctgagccc gtcgggtagg ctctcctcga 480
gttcccgttc ttcacccctt cctcaccctt cttctttcgt cacccgtecc cgacccacc 540
cgagcccgcc cctcagctg ccccgggcca tggcgtgcgc agccactctg aaaaggactc 600
tggatttcga cccgctgttg agccccggct cccgaagcg caggcgatgt gcgccattgt 660
cggcgcccac ctcggccgct gcctccccgt tgtcggcggc cgcggccacc gccgcctcct 720
tctccgctgc ggccgcctcg ccgcagaagt atctccgaat ggagccatcc cccttcggcg 780
acgtctcctc ccgcctcacc acagaacaaa ttctgtacaa cataaaacaa gagtataaac 840
gaatgcagaa gagaagacat ttagaacga gtttccaaca gacagatccg tgttgactt 900
ctgatgcaca gccacatgca tttctcctca gtggaccagc ttcaccaggg acttcatctg 960
cagcatectc accattaaaa aaagaacagc ccttatttac tctacggcag gttgggatga 1020
tctgtgaacg tttgttgaag gaacgtgaag agaaagtctg agaagaatat gaagaaatat 1080
tgaacacaaa acttgcagaa caatatgatg cgtttgtgaa gtttacgcat gatcaaataa 1140
tgcgacgata tggagaacag cctgctagct atgtttcatg aatcacgtat cctgcatttg 1200

```

tgggctgcct	tggtccctgt	tgagttgttg	caagaggtcc	caattatgac	atgcagcaat	1260
gccaataccc	cttctgtgaa	tacaggttat	ttcaagcttt	cgtcagtggc	aaccactctt	1320
aggcagcagc	aactgggtttt	ggaaatttcc	ctgatgtcag	taccacctgg	atgtggacct	1380
ttgtacctg	tattaatacc	agtggcctca	ttttgctgta	tcattacaat	ttggcttctt	1440
atattaatgt	ttgaaaagga	ttaaagctgg	tattctagaa	catgcccttc	actggttgtg	1500
taaaataaac	tgtagaatga	cacttcagat	gaagttagtg	tgattttaat	tgtgcactac	1560
aaccgagctg	taaccagtta	ctaattttag	aatgtaatcc	caggacaata	ttaagcaaat	1620
agcctgcagt	gcttctctgtg	aaatagtgaa	ggaggagggc	atctctgtat	tccaggactt	1680
cttgggggtt	cagaatgggt	ttgtatgatt	ttttttttnt	ttgtagtttt	atttattcta	1740
tcagttcttt	taacaaatgt	ttattgctgc	atcttttttt	ttccagtgtg	tcattgtttt	1800
actgcccttg	tagtactgga	atcttagttgg	aagaataaaa	catttacttc	tattttgctt	1860
gtttcttaat	gtacagatgg	ggttagtatt	tgaataaagt	tggtgtttta	aaacgtaagc	1920
atcttccagg	aatcagtgaa	gttaattttc	taagatttga	gtgctgtttc	aaaacactga	1980
gttctgatcc	taaatgcctt	cttctgctgg	gcgcgg			2016

&lt;210&gt; 38

&lt;211&gt; 2520

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2025618CB1

&lt;400&gt; 38

cccggccaag	cccgcagcgc	agggagctgt	ctgcagaggc	caggtgcgcc	tgccacgaat	60
ccccaggcac	cggtggccgc	cgcggcccga	gtagctcggc	gggtaaacat	ggccgcactg	120
acgacggttg	tggtagcggc	tgccggccacc	gcggtagccg	gggctgtggc	aggggcgggc	180
gcggccaccg	ggaccggcgt	gggagcgacg	ccagcgctc	aacagagtga	tggctgtttt	240
agtacttcag	gtggaattcg	tccttttcat	cttcagaact	ggaagcagaa	agttaatcag	300
actaagaaa	cagaatttgt	acgcacagca	gaaaaattta	aaaatcaagt	aattaacatg	360
gaaaaagata	aacacagtc	tttctacaac	caaaaaagtg	acttcagaat	tgagcatagt	420
atgctagaag	aattggaaaa	taaattgatt	cacagcagga	aaacagaaa	agcaaaaatc	480
cagcaacaat	tggccaaaat	acataataat	gtaaagaaac	ttcagcatca	attaaaagat	540
gtgaagccta	cacctgattt	tggtgagaag	ctcagagaaa	tgatggaaga	aattgaaaaat	600
gcaattaaca	cttttaagaa	agagcagagg	ttgatatatg	aagagctaat	taaagaagag	660
aagacaacta	ataatgagtt	gagtgccata	ttgacaaaaa	ttgacacatg	ggctttgggt	720
aattcagaaa	cagagaaagc	tttcagagca	atctcaagca	aagttcctgt	agacaaagta	780
acaccaagta	ctcttcagaa	agaggtacta	gattttgaaa	aattccttca	gcaaacagga	840
gggcgacaag	gtgcctggga	tgattatgat	caccagaact	ttgtaaaggt	gagaaacaaa	900
cataaaggga	agccaacatt	tatggaagaa	gttctagaac	accttcctgg	aaaaacacaa	960
gatgaagttc	aacagcatga	aaaatgggtat	caaaagtttc	tggctctaga	agaaagaaaa	1020
aaagagtc	ttcagatttg	gaaaactaaa	aagcagcaaa	aaagggagga	aattttcaag	1080
ttaaaggaaa	aggcagacaa	cacacctgtg	ctttttcata	ataaacaaga	ggataatcaa	1140
aagcaaaaag	aggaacaaa	aaagaaacag	aaattggcag	ttgaagcttg	gaagaaacag	1200
aaaagtatag	aatgtcaat	gaaatgtgct	tcccagttaa	aagaagaaga	agagaaagag	1260
aaaaaacatc	agaaagaacg	ccagcgccag	tttaagttaa	aattactact	agaaagttaa	1320
acccagcaga	agaaagaaca	ggaagaattt	ttgaggcttg	aaaaggagat	aagggaaaaag	1380
gcagaaaagg	cagaaaaaag	gaaaaatgct	gctgatgaaa	tttccagatt	tcaagaaaga	1440
gatttacata	aacttgaact	gaaaattcta	gatagacagg	caaaggaaga	tgaaaagtca	1500
caaaaacaaa	gaagactggc	aaaattaaaa	gaaaagggtg	aaaacaatgt	tagtagagat	1560
ccctctaggg	tttacaaccc	caccaaaggt	tgggaagaac	gaacccaaaa	gataggacca	1620
acaggctctg	ggccacttct	acatatccca	catagggcta	ttccaacctg	gagacaagga	1680
atacagagaa	gagtatgaga	taatcaaat	gctactcagt	tgataagaat	gttaacatac	1740
taagttatac	cagggagaga	gtgactaacc	acattcttta	aatatcaata	gcttagtcag	1800
attgattatt	gtgctatatt	gtgaattgag	aggtattaag	tttcatgagg	ctttgtcatt	1860
agtattcctg	cttctaccaa	gaaggatatt	aatatatgtg	ttggcctatt	attgatgtaa	1920
aagttatttta	aataagtttaa	tgtagaaac	attattcaat	ttaaatactg	aaaacatttc	1980
aaagagattt	tggtttttgt	atagcatagc	aaggtaaatt	ggaacaatca	tacaatgaca	2040
tttttttaaac	caaaattttg	taacttttat	aacttggagt	taagttagct	tgagtaacaa	2100
aaaggtaaa	tggtttttgt	ttagagttac	gaaatgttag	tactttttct	atgtttaaca	2160
aattggcagt	ttgtcagtta	tgacattttt	gtgtaataaa	tattttgtat	ttgtttgaag	2220
catgctttgt	tttatataga	gaatatttat	tttaaaaata	tgtctctcat	ataccctatt	2280



```

aattgtatta ttgatataat ctttttgggt tccttcagca attccaaatt ttccttcagc 2340
ctttctggat ttcacagatt tataaaatct ttgtgtcttt cacatcttcc tggctaatagc 2400
agttttcttt tctgcttctg tttgcctcaa aataggaaaa ttctttgttc tgaaacatca 2460
tctgaaataa gccagcttta aaatactgtg atttctcttg atggcactta aaatgtttta 2520

```

```

<210> 39
<211> 1036
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 3342443CB1

```

```

<400> 39
cagcagcgtc cggcgagatg aaggcgctcg gggctgtcct gcttgccctc ttgctgtgcy 60
ggcgggccagg gagagggcag acacagcagg aggaagagga agaggacgag gaccacgggc 120
cagatgacta cgacgaggaa gatgaggatg aggttgaaga ggaggagacc aacaggctcc 180
ctggtggcag gagcagagtg ctgctgcggt gctacacctg caagtccctg cccagggacg 240
agcgtgcaa cctgacgcag aactgtcac atggccagac ctgcacaacc ctcattgccc 300
acgggaacac cgagtcaggc ctctgacca ccactccac gtggtgcaca gacagctgcc 360
agcccatcac caagacgggtg gaggggaccc aggtgaccat gacctgctgc cagtccagcc 420
tgtgcaatgt cccaccctgg caaagctccc gagtccagga cccaacaggc aagggggcag 480
gcgcccccg gggcagctcc gaaactgtgg gcgcagctcc tgcctaacct ccttgccggc 540
cttgaggcaa tgggggccag gagacctga cccacggccc ctccccaccc ccaccggct 600
cacccccggc cctgccagca ctctgtctgg taccttcccc tcctgcccc gcaccagctt 660
tggagaatgg atttgagtg tcttggcgga tccagccage gcaggcccc cggccgggtt 720
gcttctcag ttcccggtg tgtccttggg gtcttcttc caccacctgt gagcagcaag 780
actgccgcac gtggggccct ggggtccagac ctgggtggc acgccccagg gccctgcagc 840
cctcacgggg ggctggggga tcgcatcagc acagccaggc agagatgata ccaccacaca 900
gctggggggc cccacaccca gtccctaccc cttaactttc tgccatgggg aatccctcca 960
tcttgaagcg gtccaaaggg gccaaccttg cccttcccc aggtcgggct tgtcagctgt 1020
tttgagggga agggggg 1036

```

```

<210> 40
<211> 1621
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2267957CB1

```

```

<400> 40
ggtgaccaca aacctatcct ctttctctca agtgaccctt ccgtacccca ccagaacatt 60
cccgggtgac ctctctccag atcttctctg tggccttcc cgcctactcc agtgacacta 120
tgacacccca ccgtgaccgg agaggcctct ggctcctgct gccgtccttg tccctgctgc 180
tttttgaggt ggccagagct ggccgagccg tggttagctg tcctgcccgc tgcctgtgcy 240
ccagcaacat cctcagctgc tccaagcagc agctgcccga tgtgccccat tccttgccca 300
gttacacagc actactggac ctcagtcaca acaacctgag ccgcctgcgg gccgagtggg 360
ccccacggc cctgacccaa ctgcactccc tgctgctgag ccacaaccac ctgaacttca 420
tctcctctga ggcttttcc ccggtaccca acctgcgcta cctggacctc tctccaacc 480
agctgcgtac actggatgag ttctgttca gtgacctgca agtactggag gtgctgctgc 540
tctacaataa ccacatcatg gcggtggacc ggtgcgcctt cgatgacatg gccagctgc 600
agaaactcta cttgagccag aaccagatct ctgccttccc tctggaactg gtcaaggaag 660
gagccaagct acccaacta acgctcctgg atctctcttc taacaagctg aagaacttgc 720
cattgcctga cctgcagaag ctgccggcct ggatcaagaa tgggctgtac ctacataaca 780
accctctgaa ctgcgactgt gagctctacc agctgttttc aactggcag tatcggcagc 840
tgagctccgt gatggacttt caagaggatc tgtactgcat gaactccaag aagctgcaca 900
atgtcttcaa cctgagtttc ctcaactgtg gcgagtacaa ggagcgtgcc tgggaggccc 960
acctgggtga caccttgatc atcaagtgtg acaccaagca gcaagggatg accaaggtgt 1020
gggtgacacc aagtaatgaa cgggtgctag atgaggtgac caatggcaca gtgagtgtgt 1080
ctaaggatgg cagtcttctt ttccagcagg tgcaggtcga ggacggtggt gtgtatacct 1140

```

```

gctatgccat gggagagact ttcaatgaga cactgtctgt ggaattgaaa gtgcacaatt 1200
tcaccttgca cggacaccat gacaccctca acacagccta taccacccta gtgggctgta 1260
tccttagtgt ggtcctgggtc ctcatatacc tatacctcac cccttgccgc tgctgggtgcc 1320
gggggtgtaga gaagccttcc agccatcaag gagacagcct cagctcttcc atgcttagta 1380
ccacacccaa ccatgatcct atggctgggtg gggacaaaga tgatgggtttt gaccggcggg 1440
tggttttccct ggaacctgct ggacctgggc aggggtcaaaa cggcaagctc aagccaggca 1500
acacctgcc agtgctgag gccacaggca agggccaacg gaggatgtcg gatccagaat 1560
cagtcagctc ggtcttctct gatacgccca ttgtgggtg agcaggatgg gttgggtggg 1620
a 1621

```

&lt;210&gt; 41

&lt;211&gt; 3562

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7480277CB1

&lt;400&gt; 41

```

ccgctctggg cccccgcagg ccaaggccgc cgggcccggg cggggacgac caacttgggg 60
cgcgccgtag ccccgctctc ccgagagctc ggagcccggg agggctacgg ccgcccgcag 120
acggcgggag aggagcgagg cgagcggagg cggcgagcgg cggccgcgccc gcagccccgg 180
cctgggaact ttctctctt ccccggtttt tggggccctt ctgacctgga attgctctcc 240
agattccccg ggggcgcagg gctgctattc ttccccggg tttatcgggc gctcgggctaa 300
cttcacggac ccggggaccc gcggcgctcg tccctcgccc gaacccagcc cgcgctgctc 360
cccgatcag gagggccggg cccggggctg ctgcgcgcc gcgagtgcct tcagcccggc 420
cccctggagt cgggcccgtg agccacggc agcggccgca ggactggaaa cagcagattg 480
attaactcga gcggagcccc ggccctcccc actccgctcc gctgaggggc ggccccagtg 540
cggggaaacg acaagtttgt cagtcgtccg tggcctgttg gatcgaagcg ccgcctccgc 600
cgccgagagg tccccggcgc ctacgatccc gcgcggacgg ccctgggtac ccggggcggc 660
tcggcggcgg ggctcctcgg gtccggggcgc tggtctgtgt gccgggcgcg ccgaggcacc 720
cggggctggg ccagcgcgcc ctgcgtcccc agcggggcag cggccccgcc ggaggagaaa 780
cacgggtcgc cgccacctcc gcctcttcag tctcctggtc ttctgcgcg ctctctctct 840
cacctctcag ggaaaggggg ggacataggg gcgtcgcggg gccccggcga atgcgcccc 900
cgccgcctct cgggctgcgc cgctcgcgg ggatgaagca ccggccgtga agatggaggt 960
gacctgcctt ctacttctgg cgctgatccc ctccactgc cggggacaag gactctacgc 1020
tcagcccag gcgcagatcg tgcatgcggg ccaggcatgt gtggtgaaag aggacaatat 1080
cagcgagcgt gtctacacca tccgggaggg ggacaccctc atgctgcagt gccttgtaac 1140
agggcaccct cgaccccagg tacggtggac caagacggca ggtagcgcct cggacaagtt 1200
ccaggagaca tcggtgttca acgagacgct gcgcacgag cgtattgcac gcacgcaggg 1260
cggccgctac tactgcaagg ctgagaacgg cgtgggggtg ccggccatca agtccatccg 1320
cgtggacgtg cagtcctatga agaacgctac attccagatc actcctgacg tgatcaaaga 1380
gagtgagaac atccagctgg gccaggacct gaagctatcg tgccacgtgg atgcagtgcc 1440
ccaggagaag gtgacctacc agtggttcaa gaatggcaag ccggcacgca tgtccaagcg 1500
gctgctgggt acccgcaatg atcctgagct gccgcagtc accagcagcc tagagctcat 1560
tgacctgcac ttcagtgact atggcaccta cctgtgcatg gcttctttcc caggggcacc 1620
cgtgcccgac ctacagctcg aggtcaacat ctccctctgag acagtgcgc ccaccatcag 1680
tgtgcccaag ggtaggggccg ttggtgaccgt gcgcgagggg tgcctgccc agctgcaatg 1740
cgaggtgcgg ggcaagccgc ggccgccagt gctctgggtc cgctgggaca aggaggctgc 1800
actgctgccc tcggggctgc ccctggagga gactccggac gggaagctgc ggctggagcg 1860
agtgagccga gacatgagcg ggacctaccg ctgcccagag gcccgctata atggcttcaa 1920
cgtgcgcccc cgtgaggccc aggtgcagct gaacgtgcag ttcccgcgg aggtggagcc 1980
cagttcccag gacgtgcgcc aggcgtggg ccggcccgtg ctccctgcgt gctcgtgct 2040
gcgaggcagc cccagcgca tcgcctcgcc tgtgtggcgt ttcaaagggc agctgctgcc 2100
gccgcccct gttgttccc cgcccgccga ggcccggtat cacgcggagc tgcgctcga 2160
cgccgtaact cgcgacagca gcggcagcta cgagtgcagc gtctccaacg atgtgggtc 2220
ggctgctgc ctcttccagg tctccgcca agcctacag ccggagtttt acttcgacac 2280
ccccaacccc acccgagcc acaagctgc tctaactgc tgcagtggag tgcagtggag 2340
tcagagggag cccgacgtg tcgacctgt gctcaactac agactcagca tccgccagtt 2400
gaaccagcac aatgcggtgg tcaaggccat cccgggtccg cgtgtggaga aggggcagct 2460
gctggagtac atcctgaccg atctccgtgt gcccacagc tatgaggtcc gcctcacacc 2520
ctataccacc ttcggggctg gtgacatggc ctcccgcac atccactaca cagagcgcca 2580

```

```

gatccgctgg cccccagtcc tggctctgag gaccctgtcc tctgggtccca agcagggtat 2640
cctctgcaga gccccacacc tcagttctga cttgggttcc ccgcttgctt tctcagccat 2700
caactctccg aacctttcag acaacacctg ccactttgag gatgagaaga tctgtggcta 2760
taccaggac ctgacagaca actttgactg gacgcggcag aatgccctca cccagaaccc 2820
caaacgctcc cccaacactg gtccccccac cgacataagt ggcacccctg agggctacta 2880
catgttcac gagacatcga ggccctcgga gctgggggac cgtgcaaggc tagtgagtcc 2940
cctctacaat gccagcgcca agttctactg tgtctccttc ttctaccaca tgtacgggaa 3000
acacatcggc tccctcaacc tcctggtgcg gtcccggaac aaaggggctc tggacacgca 3060
cgcctggtct ctcagtggca ataagggcaa tgtgtggcag caggcccatg tgcccatcag 3120
ccccagtggg cccttcacga ttatttttga ggggggttca ggcccggtc acctggggga 3180
tattgccata gatgacgtca cactgaagaa gggggagtgt ccccggaagc agacggatcc 3240
caataaaggc gcaagacggg aaggagctgc ctgcgatggc ctgaaattcc acctttcatc 3300
ccctatggat gacggagagc ttacagatga ccctattgaa tgcaagcacc tttggatcca 3360
tagagtggac agtaaaggct ctcagtacat gttggctgag ctgaactgca tacatgtggc 3420
ccccaggttc ctggtcttta tggacgaagg gcacaagggt ggtgaaaagg actccggggg 3480
ccaggtgctg tatagcagct tatggaagtc tcagctgggc tatcctgccc ttgggagcac 3540
agacaggctc ctagggtgct ga

```

<210> 42  
 <211> 899  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3450647CB1

```

<400> 42
caggaattcg gcacgaggga agtctgaaat caaggggtca gcaggggttg tttcttcagc 60
agagtctacg gaagaatccc tcccacatct ccctcctcgc tttgggtggc ttctggccat 120
ccctgctgtg ccttgacttg tagatgtcac tcccgtttct gcttgcatct ttgcttggcc 180
ttcttcctta cgtctgtgtg tctcctcttc ggtctcttct aaggacatgt gtcgttagat 240
ttatggccca cctagtgcca ggacaatctc atcttgagat ccttaatcta attacatttg 300
caaagtcctt tttcgcaata aggtcacgtt cacagggtcca gagattagga cttaaacata 360
tcttttcttg gggctggggg ggacactatt caacccctg cagtgcctg ggggctctca 420
tctttatttt cctcattagt aaaatgggct catgttatct gctttacagg attgctgtaa 480
atattaaaga aaataatata ttcctggccg agcacagtgg ctcatgcctg ccagtctcag 540
cctcccaaaa tgccaggatt acaggcatga gccaccatgc ccggcccttg gtaataacta 600
ttcttaatgt gttttatcac agtttaact cttacctcct ctgtcgtgct cccacaccct 660
ggatcatgaga tatatttgaa atgggagcca cgggcaaggg gtctttgcat gtctagccta 720
gtgtctcgcg catagtaaaa ggataaagaa tatttgtttt gactgtacct gagtagttca 780
catggaaatt actgaagccc tgtggtgctg atgtgttaag taatactgct gccacgactg 840
tttatcaaac atgtatgggg tgagatatta cttttaccca cgagttgctc caaaaaaaaa 899

```

<210> 43  
 <211> 2330  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2053428CB1

```

<400> 43
ttcgtgggta gttttttttg ccagcttcta taccgtcaca cagagggtctg cacctgttgt 60
ccactgctaa cggcagaggc tccaggtacc ccgacctgac gcgcacaaaa cttgggtgtg 120
taagaatggc cccaacgcgc tccacaagg acacacgctt gtcccccgga atttttggcg 180
ggcaacctca taatctgtcg caactcaaca tagtaccacc ccggtcgttt gtggcaagac 240
cggagacaag ttaaggtgcc actcaggccg ggtccaaaca cgcaagttga agagtctccc 300
gcatagtgag ttcgcaatcg tctcgacatg acacggggtc tctcattca gatcaaaagg 360
cgcgcatatt tcccccttg gagtggaaca accggatcct cgtgccttgg ttaactggaa 420
acggaatcgg tcgtcgtctg ctcccggcaa tcgcaagcc ttctctcta gccccgtaca 480
caatagttcc gtctcgctag cgccaataa gtctggacga ccgcaagggg taagcaagcc 540

```

```

ggccggatga gaaagcatag agaccggaaa tgtgcctgtt tcttcctgtc ctaagttcgg 600
agtcagcgcc ccttgtgggc cggaaggga gtagcgttgt tgctgggaag atggcgaccg 660
cggcgactat cccatcggtg gccacggcca cagcagcggc tctcggcgag gtggaggatg 720
aagggtcctt ggcgtcgctg ttccgggacc gcttccccga ggcccagtgg cgcgagcggc 780
ccgatgtggg ccgtaccttc cgggagtga gcggctcggg gctggagcgg ctgcggcgcg 840
agccccgagc cctggcggag gagcgggcgc agctgctgca gcagacgcgc gacttggcct 900
tcgctaacta caagaccttc atccgcggcg ccgagtgcac cgagcgcac caccgcctgt 960
ttggcgacgt ggaggcgtcg ctccggccgc tgctcgaccg tttgccagc ttccagcaga 1020
gctcgaggaa ctttgtgaag gaagccgagg agatcagctc caaccgcgg atgaatagcc 1080
tgaccctaaa ccggcacaca gaaattttgg aaatactgga gattcctcag ctcatggaca 1140
cctgtgtccg gaacagttat tatgaagagg ccctggagct tgcagcctac gtacgcccag 1200
tggagaggaa atactcttcc atccctgtca tccagggcac cgtgaacgaa gtgcgccagt 1260
ccatgcagct gatgctgagc cagctgatec agcaactgag gaccaacatc cagcttcctg 1320
cctgcctccg tgtcattggc tacctgcggc gcatggacgt cttcactgag gctgagttga 1380
gggtgaagtt tcttcaggcc cgagatgctt ggctccggtc cactcctgact gccattccta 1440
atgatgatcc ctatttccat attacaaaaa ccacgcaggc ctcccgctgc catctctttg 1500
atatcatcac ccagtaccgt gccatcttct cagacgagga cccactgctg ccccctgcca 1560
tgggtgagca cactgtgaat gagagtgcga tcttccatgg ctgggtgcta cagaaggtct 1620
cacaattcct gcaggtgctg gagaccgacc tttaccgggg cataggcggc cacttggaact 1680
ctctgctggg ccagtgcagc tactttgggc tgccttcag ccgggtggga gctgatttcc 1740
ggggtcagtt ggctcctgtt ttccagcggg tggccatcag cactttccag aaagcaattc 1800
aggaaacagt ggagaaattc caggaaagaa tgaactccta catgctcatc tcggctccag 1860
ccatcctggg caccagtaac atgcctgctg ctgtgccagc caccagccg gggacgctgc 1920
agccacccat ggtgctccta gatttcccac ccctgcctg ctttctcaac aatattctgg 1980
ttgccttcaa tgatctgcgc ctctgctgcc ctgtggccct ggcgaggat gtgactgggg 2040
ccttgaaga tgcccttggc aaggttaact aaataatcct ggccctccat cgcgctgaag 2100
aggctgcctt cagcagcggg gagcaagagc tctttgtcca gttctgact gtcttcctgg 2160
aagaccttgt tccgtattta aatcgctgtc tccaagtcct ttttccacca gctcagatag 2220
cacagacttt aggttaagaga atgaaaattc tgtaaaactgc cttactgatg tgtaattcac 2280
atactaaata attcatccat ttgtgtccac caaacaacaa agggggccgc 2330

```

&lt;210&gt; 44

&lt;211&gt; 1755

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7503614CB1

&lt;400&gt; 44

```

ctacgagggg ggggaagcgg ttggacgtgt tcgcttgggt tcctgctgcg gcagctacct 60
cgcaatctct ctgcatcgat cgccgctcgc aagctactga ccgtactcgg gcgtattagg 120
agccgcgttc cagcctcaca ccccacggtg ctgttttcga cttcagaaag gatctagcct 180
cagcacagaa gcgcctcagg cgcggcgcaa agctcgagcg gacggcgggg gcggccggag 240
cctctctcgg gggagccgcg cctgaggagg cgaagaacc cccctgacgc gactggcgtg 300
tgcttctgcc cgccaccgcc cctcccgcct tcaccggggc cgtcccctggc cactgcccct 360
gccgcggagg cagcggcggc agcggctctc ctttccacag ccggcgctcc gcgaccgcct 420
tggctcctga gcccgctcgg taggctctcc tcgagttccc gctcttcacc ccttccctca 480
ccctcttctt tcgtcaccgc tcccgcaccc caccgcagcc cggcgctcca gctgcccccg 540
gccatggcgt gcggagccac tctgaaaagg actctggatt tcgaccgcgt gttgagcccc 600
gcgtccccga agcgcaggcg atgtgcgcca ttgtcggcgc ccacctcggc cgtgcctccc 660
ccgttgtcgg cggccgcggc caccgcgcgc tcttctcgg ctgcggcgcc ctcgcccag 720
aagtatctcc gaatggagcc atcccccttc ggcgacgtct cctccgcct caccacagaa 780
caaattctgt acaacataaa acaagagtat aaacgaatgc agaagagaag acatttagaa 840
acgagtttcc aacagacaga tccgtgttgt acttctgatg cacagccaca tgcatttctc 900
ctcagtgga cagcttcacc agggacttca tctgcagcat cctcaccatt gttccttgtt 960
gagttgttgc aagaggtccc aattatgaca tgcagcaatg ccaatacccc ttctgtgaat 1020
acaggttatt tcaagctttc gtcagtggca accactctta ggagcagca actggttttg 1080
gaaatttccc tgatgtcagt accacctgga tgtggacctt tgctacctgt attaatacca 1140
gtggcctcat tttgctgtat cattacaatt tggcttctta tattaatgtt tgaaaaggat 1200
taaaagctgg attctagaac atgcccttca ctggttgtgt aaataaaact gtagaatgac 1260
acttcagatg aagttagtgt gattttaatt gtgcactaca accgagctgt aaccagttac 1320

```

```

taat t t t t t a g a t g t a a t c c c a g g a c a a t a t t a a g c a a a t a g c c t g c a g t g c t t c c t g t g a 1380
a a t a g t g a a g g a g g a g g g c a t t t c t g t a t t c c a g g a c t t c t t g g g g t t t c a g a a t g g g t t 1440
t g t a t g a t t t t t t t t t t t t t g t a g t t t t a t t t a t t c t a t c a g t c t t t t t a a c a a a t g t t 1500
t a t t g c t g c a t t t t t t t t t t t t c c a g t g t a t c a t t g t t t t a c t g c c c t t g t a g t a c t g g a a 1560
t t t a g t t g g a a g a a t a a a a c a t t t a c t t c t a t t t t g c t t g t t t c t t a a t g t a c a g a t g g g 1620
g t t a g t a t t t g a a t a a a g t t g g t g t t t t a a a a c g t a a g c a t t t t c c a g g a a t c a g t g a a g 1680
t t a a t t t t c t a a g a t t t g a g t g c t g t t t c a a a a c a c t g a g t t c t g a t t c t a a a t g c c t t c 1740
t t c t g c t g g g c g c g g 1755

```

&lt;210&gt; 45

&lt;211&gt; 2427

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7503456CB1

&lt;400&gt; 45

```

a t c c c g g c c a a g c c c g c a g c g c a g g g a g c t g t c t g c a g a g g c c a g g g g t g c g c c t g c c a c g 60
a a t c c c c a g g c a c c g g t g g c c g c c g c g g c c c a g t a g c t c g g c g g g t a a a c a t g g c c g c a 120
c t g a c g a c g g t t g t g g t a g c g g c t g c g g c c a c c g c g g t a g c c g g g g c t g t g g c a g g g g c g 180
g g c g c g g c c a c c g g g a c c g g c g t g g g a g c g a c g c c a g c g c t c a a c a g a g t g a t g g c t g t 240
t t t a g t a c t t c a g g t g g a a t t c g t c c t t t t c a t c t t c a g a c t g g a a g c a g a a g t t a a t 300
c a g a c t a a g a a a g c a g a a t t t g t a c g c a c a g c a g a a a a a t t a a a a a t c a a g t a a t t a a c 360
a t g g a a a a a g a t a a a c a c a g t c a t t t t c t a c a a c c a a a a a g t g a c t t c a g a a t t g a g c a t 420
a g t a t g c t a g a a g a a t t g g a a a t a a a t t g a a t t c a c a g c a g g a a a c a g a a g a g c a a a a 480
a t c c a g c a a c a a t t g g c c a a a t a c a t a a t a a t g t a a a g a a a c t t c a g c a t c a a t t a a a a 540
g a t g t g a a g c c t a c a c c t g a t t t t g t t g a g a a g c t c a g a g a a a t g a t g g a a g a a a t t g a a 600
a a t g c a a t t a a c a c t t t t t a a a g a g a g c a g a g g t t g a t a t a t g a a g a g c t a a t t a a a g a a 660
g a g a a g a c a a c t a a t a a t g a g t t g a g t g c c a t a t c a a g a a a a t t g a c a c a t g g g g c t t t g 720
g g t a a t t c a g a a c a g a g a a a g c t t t c a g a g c a a t c t c a a g c a a a g t t c c t g t a g a c a a a 780
g t a a c a c c a a g t a c t c t t c c a g a a g a g g t a c t a g a t t t t g a a a a t t c c t c a g c a a a c a 840
g g a g g g c g a c a a g g t g c c t g g a t g a t t a t g a t c a c c a g a a c t t t g t a a a g g t g a g a a a c 900
a a a c a t a a a g g a g a g c c a a c a t t t a t g g a a g a a g t c t a g a g t t c t a a c a c c t t c c t g g a a a a c a 960
c a a g a t g a a g t t c a a c a g c a t g a a a a a t g g t a t c a a a a g t t t c t g g c t c t a g a a g a a g a 1020
a a a a a a g a g t c a a t c a g a t t g g a a a a c t a a a a g c a g c a a a a a g g g a g g a a a a t t t t c 1080
a a g t t a a a g g a a a a g g c a g a c a c c t g t g c t t t t t c a t a a t a a c a a g a g g a t a a t 1140
c a a a g c a a a a a g a g g a a c a a g a a a g a a a c a g a a a t t g g c a g t t g a a g c t t g g a a g a a a 1200
c a g a a a a g t a t a g a a t g t c a a t g a a t g t g c t t c c c a g t t a a a a g a a g a a g a g a a 1260
g a g a a a a a a c a t c a g a a a g a a c g c c a g c g c c a g t t t a a g t t a a a a t t a c t a c t a g a a a g t 1320
t a t a c c c a g c a g a a g a a g a a c a g g a a g a a t t t t g a g g c t g a a a a g g a g a t a a g g g a a 1380
a a g g c a g a a a a g g c a g a a a a a a g g a a a a a t g c t g c t g a t g a a a t t c c a g a t t t c a a g a a 1440
a g a g t t g a a a a c a a t g t t a g t a g a g a t c c c t c t a g g c t t t a c a a c c c a c a a a g g t t g g 1500
g a a g a a c g a a c c a a a a a g a t a g g a c c a a c a g g c t c t g g g c c a c t t c t a c a t a t c c c a c a t 1560
a g g g c t a t t c c a a c c t g g a g a c a a g g a a t a c a g a g a a g t a t g a g a t a a t c a a a t t g c t 1620
a c t c a g t t g a t a a g a a t g t t a a c a t a c t a a g t t a t a c c a g g t a t a t t g t g t a t a t t g t g 1680
t t c t t t a a a t a t c a a t a g c t t a g t c a g a t t g a t t a t t g t g c t a t a t t g t g a a t t g a g a g g 1740
t a t t a a g t t t c a t g a g g c t t t g t c a t t a g t a t t c c t g c t t c t a c c a a g a a g g t a t t t a a t 1800
a t a t g t g t t g c c t a t t a t t g a t g t a a a a g t t a t t t a a a t a a g t t a a t g t t a g a a c a t t 1860
a t t c a a t t t a a a t a c t g a a a a c a t t t t c a a a g a g a t t t t g t t t t g t t a t a g c a t a g c a a a 1920
g t a a a t t g g a a c a a t c a t a c a a t g a c a t t t t t t a a c c a a a a t t t t g t a a c t t t t a t a a c 1980
t t g g a g t t a a g t t a g c t t g a g t a a c a a a a a g g t a a a g t g g t t t t g t t t a g a g t t a c g a a 2040
a t g t t a g t a c t t t t t c t a t g t t t a a c a a a t t g g c a g t t t g t c a g t t a t g a c a t t t t t g t g 2100
t a a t a a a t a t t t t g t a t t t g t t t g a a g c a t g c t t t g t t t t a t a t a g a g a a t a t t t a t t t t 2160
a a a a a t a t g t c t c t c a t a t a c c c t a t t a a t t g t a t t a t t g a t a t a a t c t t t t g g t t t c c 2220
t t c a g c a a t t c c a a a t t t t c c t t c a g c c t t c t g g a t t t c a c a g a t t t a t a a a a t c t t t g 2280
t g t c t t t c a c a t c t c c t g g c t a a t g c a g t g c t t c t t t c t g c t c t g t t t g c c t c a a a a t 2340
a g g a a a a t t c t t t g t t c t g a a c a t c a t c t g a a a t a a g c c a g c t t t a a a a t a c t g t g a t t 2400
t c t c t t g a t g g c a c t t a a a a t g t t t t a 2427

```

&lt;210&gt; 46

&lt;211&gt; 1685

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7503459CB1

&lt;400&gt; 46.

```

tactatttcg ctatcttctc aaacatgaca cgggtctcct cttcagtcaa aggccgcatt 60
tttccccctg ggttggaaca ccggaacctc gtgccttggg aactggaaac ggaagtcggg 120
cgctcgctgc tccccggcaa tcccaaagcc ttctctctta gccccgtacc aatagttcgt 180
ctcgctagcg cccaatagtc tggâcgcgacg caggggaaag caagccggcc ggatgagaaa 240
gcatagagac cggaaatgtg cctgtttctt cctgtcctaa gttcggagtc agcgcctctt 300
gtggtccgga agggaagtga cgttggtgct gggaagatgg cgaccgcggc gactatccca 360
tcggtagcca cggccacagc agcggctctc ggcgaggtgg aggatgaagg gctcctggcg 420
tcgctgttcc gggaccgctt ccccagggcc cagtggcgcg agcggcccga tgtgggccgc 480
tacctccggg agttgagcgg ctcggggctg gagcggctgc ggcgcgagcc cgagcgctg 540
gcggaggagc gggcgcgagc gctgcagcag acgcgcgact tggccttcgc taactacaag 600
accttcatcc gcggcgccga gtgcaccgag cgcattccacc gcctgtttgg cgacgtggag 660
gcgtcgctcg gccgcctgct cgaccgtttg cccagcttcc agcagagctg caggaaactt 720
gtgaaggaag ccgaggagat cagctccaac cgcgggatga atagcctgac cctaaaccgg 780
cacacagaaa ttttggaat actggagatt cctcagctca tggacacctg tgtccggaac 840
agttattatg aagaggccct ggagcttgca gcctacgtac gccgactgga gaggaaâtac 900
tcttccatcc ctgtcatcca gggcatcgtg aacgaagtgc gccagtcâat gcagctgatg 960
ctgagccagc tgatccagca actgaggacc aacatccagc ttctgcctg cctccgtgct 1020
attggctacc tgcggcgcat ggacgtcttc actgaggctg agttgagggt gaagtttctt 1080
caggcccag agcttggct ccggtccatc ctgttttcca gcgggtggcc atcagcactt 1140
tccagaaagc aattcaggaa acagtggaga aattccagga agaaatgaac tcctacatgc 1200
tcatctcggc tccagccatc ctgggcacca gtaacatgcc tgctgctgtg ccagccaccc 1260
agccggggac gctgcagcca cccatgggtg tcctagattt cccaccctc gcctgcttct 1320
tcaacaatat tctggttgcc ttcaatgatc tgcgcctctg ctgccctgtg gccctggcgc 1380
aggatgtgac tggggccttg gaagatgcc ttgccaaggt aactaaaata atcctggcct 1440
tccatcgcg tgaagaggct gccttcagca gcggggagca agagctctt gtccagttct 1500
gactgtctt cctggaagac cttgttccgt atttaaactg ctgtctccaa gtccttttct 1560
caccagctca gatagcacag actttaggta agagaatgaa aattctgtaa actgccttac 1620
tgatgtgtaa ttcacatact aaataattca tccatttgtg tccaccaaac aaaaaagggg 1680
gccgc 1685

```

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.